



# Conception et évaluation de nouvelles techniques d'interaction dans le contexte de la télévision interactive

Dong-Bach Vo

## ► To cite this version:

Dong-Bach Vo. Conception et évaluation de nouvelles techniques d'interaction dans le contexte de la télévision interactive. Interface homme-machine [cs.HC]. Télécom ParisTech, 2013. Français. NNT : 2013ENST0053 . tel-01233287

**HAL Id: tel-01233287**

**<https://pastel.archives-ouvertes.fr/tel-01233287>**

Submitted on 24 Nov 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



EDITE - ED 130

## Doctorat ParisTech

### T H È S E

pour obtenir le grade de docteur délivré par

**TELECOM ParisTech**

**Spécialité « Informatique et Réseaux »**

*présentée et soutenue publiquement par*

**Dong-Bach VO**

le 24 septembre 2013

## **Nouvelles Techniques d'Interaction Gestuelle pour la Télévision Interactive**

Directeur de thèse : **Eric LECOLINET**

Directeur de thèse : **Yves GUIARD**

### **Jury**

**M. Stéphane HUOT**, Maître de Conférences H.D.R., UMR 8623, Université Paris Sud XI

**M. Nicolas ROUSSEL**, Directeur de Recherche, UMR 8022, Université Lille I

**M. Renaud BLANCH**, Maître de Conférences, UMR 5217, Université Grenoble I

**M. Bertrand DAVID**, Professeur des Universités, UMR 5205, Ecole Centrale de Lyon

Rapporteur

Rapporteur

Examineur

Examineur

**TELECOM ParisTech**

école de l'Institut Mines-Télécom - membre de ParisTech

46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - [www.telecom-paristech.fr](http://www.telecom-paristech.fr)



---

## ABSTRACT

---

Since the 40's, television has never stopped being popularized and offering new services to the viewers. These interactive services make viewers more engaged in television activities. Unlike the use of a computer, these viewers seek information, rest, entertain themselves, sitting comfortably on a sofa. Not wanting to use a keyboard and mouse, they interact on a remote screen with a remote control and applications which are often poorly designed for this environment. This device and the current interaction techniques associated with it are struggling to meet viewers' expectations.

To address this problem, the work of this thesis explores the possibilities offered by the gestural modality for designing new interaction techniques for interactive television, taking into account its context of use.

More specifically, in a first step, we present the specific context of the television usage. Then, we propose a literature review of research trying to improve the remote control. Finally we focus on gestural interaction. To guide the design of interaction techniques based on gestural modality, we introduce a taxonomy that attempts to unify gesture interaction constrained by a surface and hand-free gestures.

Therefore, we propose various techniques for gestural interaction in two scopes of research : gestural instrumented interaction techniques and hand-free gestural interaction. We first developed techniques based on the familiar use of the remote control and the combination of gestures that do not monopolize visual attention of the viewers on the remote. The results of the comparative study of these techniques show that the combination of gestures and buttons of the remote control is very efficient and can increase the interactional bandwidth while keeping the interaction techniques familiar from the viewers. Concerning hand-free interaction, we explored the possibility of performing gestures on the surface of the belly without glancing at it. The results of our studies show that users have multiple mental spatial representations of orientation when they perform these gestures. Building upon these results, we propose an analysis of gestures that do not demand significant cognitive effort.



---

## RÉSUMÉ

---

Depuis son apparition dans les années 40, la télévision n'a cessé de se populariser et d'évoluer en proposant de nouveaux services. Ces services de plus en plus interactifs rendent les téléspectateurs plus engagés dans l'activité télévisuelle. Contrairement à l'usage d'un ordinateur, ces téléspectateurs cherchent à s'informer, à se reposer, ou se divertir, confortablement assis dans leur canapé. Ne souhaitant pas utiliser de clavier et de souris, ils interagissent sur un écran distant avec une télécommande et des applications souvent mal conçues pour cet environnement. Ce dispositif et les techniques d'interaction actuelles qui lui sont associées peinent à répondre correctement à leurs attentes.

Afin de répondre à cette problématique, les travaux de cette thèse explorent les possibilités offertes par la modalité gestuelle pour concevoir de nouvelles techniques d'interaction pour la télévision interactive en tenant compte de son contexte d'usage.

Plus précisément, dans un premier temps, nous présentons le contexte singulier de l'activité télévisuelle. Puis, nous proposons un espace de caractérisation des travaux de la littérature cherchant à améliorer la télécommande pour, finalement, nous focaliser sur l'interaction gestuelle. Afin de guider la conception de techniques d'interaction basées sur la modalité gestuelle, nous introduisons un espace de caractérisation qui tente d'unifier l'interaction gestuelle contrainte par une surface, mains libres, et instrumentée ou non.

Enfin nous proposons diverses techniques d'interaction gestuelle selon deux axes de recherche : les techniques d'interaction gestuelle instrumentées et les techniques d'interaction gestuelles mains libres. Nous avons d'abord conçu des techniques basées sur l'usage familier de la télécommande et la combinaison de gestes qui ne sollicitent pas le regard des téléspectateurs sur la télécommande. Les résultats de l'étude comparative de ces techniques montrent que la combinaison de gestes et la pression des boutons de la télécommande est très performante et permet d'augmenter la bande passante interactionnelle tout en conservant les techniques d'interaction familières des téléspectateurs. Pour l'interaction mains libres, nous avons exploré la possibilité de réaliser des gestes, sans regarder, sur la surface du ventre. Les résultats de nos études révèlent que les utilisateurs possèdent plusieurs représentations spa-

tiales mentales de l'orientation lorsqu'ils réalisent ces gestes. En nous servant de ces résultats, nous proposons une analyse des gestes qui ne demandent pas d'effort cognitif important.



*Soyons reconnaissants aux personnes qui nous donnent du bonheur ;  
elles sont les charmants jardiniers par qui nos âmes sont fleuries.*

Marcel Proust

---

## REMERCIEMENTS

---

L'aboutissement de cette thèse n'aurait pu être possible sans l'aide et le soutien de nombreuses personnes à qui je tiens à exprimer ma gratitude.

En premier lieu, je tiens à remercier Eric Lecolinet et Yves Guiard, mes directeurs de thèse. Tout au long de ce voyage, ils n'ont eu cesse de me motiver quand bien même il était difficile de penser que je pourrais finalement mener à bien ce travail. Je les remercie de m'avoir donné la chance d'intégrer cette équipe dans laquelle j'ai rencontré des personnes formidables et de m'avoir fait confiance jusqu'au bout. Je les remercie également pour tous les conseils avisés qu'ils ont su m'apporter pour me guider et les connaissances qu'ils m'ont prodigué pour me former à la recherche.

Je souhaite remercier mes rapporteurs Stéphane Huot et Nicolas Roussel. Je suis très honoré pour le temps qu'ils ont consacré à l'évaluation approfondie de mon travail de thèse. Ce sont deux chercheurs que je respecte beaucoup et qui m'inspirent au quotidien dans mes propres travaux. Je voudrais également remercier mes examinateurs Renaud Blanch et Bertrand David qui m'ont honoré de leur présence et éclairé par leurs remarques constructives qui permettront d'améliorer encore mon travail de recherche.

Je désire également remercier les membres permanents du département INFRES et plus particulièrement Patrick Bush, Ada Diaconescu, Jean-Louis Dessales, Jean Leneutre, Elie Najm, Thomas Robert, Ahmed Serhouchni et Sylvie Vignes. J'ai eu l'occasion de discuter avec eux à maintes reprises sur divers sujets qui ont alimenté mes activités de recherche et mes activités d'enseignement de près ou de loin.

Je remercie Hayette Sousou et Céline Bizart pour leur aide dans toutes les démarches administratives concernant la vie au laboratoire. Je remercie particulièrement Florence Besnard qui m'a accompagné dans toutes les procédures administratives liées à la vie doctorale. Venir chercher quelques friandises dans son bureau a toujours été un grand plaisir.

Je remercie les nombreux doctorants de tous les départements avec qui j'ai partagé d'agréables moments pendant mon séjour à Télécom Paristech. Ces moments ont été tous autant nécessaires que les moments d'isolation pour l'aboutissement de cette thèse.

Tout au long de cette thèse, j'ai rencontré des gens formidables qui sont devenus des amis proches. Leur immense soutien, leurs conseils judicieux et leur précieuse amitié ont été indispensables à la réussite de cette aventure. Je remercie Mathias Baglioni, Gilles Bailly, Dragutin Brezak, Hind Gacem, Sébastien Gardoll, Sylvain Malacria, Halla Olafsdottir, Simon Perrault, Thomas Pietrzak, Anne Roudaut, Quentin Roy, Marcos Serrano et Julie Wagner pour tous ces cafés, bières, soirées, journées à Disneyland, mariages, voyages, discussions, délires, tous ces moments qui nous ont rapproché un peu plus à chaque fois. Vous ne pouvez imaginer ce que tout cela représente pour moi. Je ne sais pas où l'avenir me mènera, mais sachez que je ne vous oublierai jamais.

Sans cette passion qui m'anime pour l'Interaction-Homme Machine et la recherche, ces remerciements n'existeraient pas. Je remercie Philippe Truillet, Emmanuel Dubois, Mathieu Raynal et Yannick Jestin pour m'avoir donné l'opportunité de découvrir ce domaine passionnant. Je remercie également Marco Winckler pour m'avoir donné la chance de découvrir la recherche. Par la même occasion, j'aimerai également remercier Philippe Palanque. Il reste une source d'inspiration dans mes discussions et activités de recherche encore aujourd'hui. Après ces années de doctorat, je comprends mieux certaines choses et les discussions que nous avons partagées prennent maintenant toutes leur sens.

Vivre à Paris n'est pas toujours facile. Je remercie mon cousin Sahn Nguyen et son épouse Phuong qui ont contribué à rendre mon séjour à Paris très agréable. De même, je remercie Mi pour ses plats qui me régaleront toujours autant.

Enfin, je voudrais conclure en remerciant ma famille proche. Je remercie mon frère Vinh et Désiré Bonhoure pour le soutien et l'amour qu'ils me témoignent au quotidien. Je remercie également mes parents Huu Vinh et Mong Quyen. Je ne pourrai jamais exprimer l'immense gratitude que j'ai à leur égard. Ils m'ont soutenu durant toutes ces années et m'ont permis d'arriver si loin. Je remercie tout particulièrement ma mère qui a toujours oeuvré pour nous offrir, à mon frère et moi-même, un avenir lumineux, au coût de nombreux sacrifices. Aujourd'hui, j'ai la chance d'avoir trouvé un domaine

passionnant dans lequel je pense pouvoir m'épanouir professionnellement.

Merci Maman.



---

## TABLE DES MATIÈRES

---

1	INTRODUCTION	1
1.1	Contexte général . . . . .	1
1.2	Organisation du mémoire . . . . .	2
2	CONTEXTE	5
2.1	De la télévision à la télévision interactive . . . . .	5
2.2	La télévision interactive . . . . .	6
2.3	Une activité avec différentes caractéristiques . . . . .	11
2.3.1	Une activité sociale . . . . .	11
2.3.2	Une activité à différents degrés d'engagement . . . . .	13
2.3.3	Une activité de divertissement . . . . .	14
2.4	Conclusion . . . . .	15
3	ESPACE DE CARACTÉRISATION DES TÉLÉCOMMANDES	17
3.1	La télécommande traditionnelle . . . . .	18
3.2	Types de télécommandes . . . . .	20
3.2.1	Télécommandes dédiées . . . . .	21
3.2.2	Télécommandes non dédiées . . . . .	29
3.3	Dimension interactionnelle . . . . .	33
3.3.1	Entrée . . . . .	33
3.3.2	Sortie . . . . .	42
3.4	Conclusion . . . . .	45
4	ESPACE DE CARACTÉRISATION DU GESTE DE LA MAIN POUR LA TÉLÉVISION INTERACTIVE	49
4.1	Taxonomies existantes en interaction homme-machine . . . . .	49
4.2	Espace de caractérisation proposé . . . . .	57
4.2.1	Type . . . . .	60
4.2.2	Fonction . . . . .	60
4.2.3	Anatomie gestuelle . . . . .	63
4.2.4	Contexte . . . . .	70
4.2.5	Conclusion . . . . .	71
5	TECHNIQUE D'INTERACTION GESTUELLE INSTRUMENTÉE POUR LA TÉLÉVISION	73
5.1	Introduction . . . . .	73
5.2	Type du mouvement du dispositif . . . . .	73
5.3	Amplitude du mouvement du dispositif . . . . .	77



5.4	Correspondance entre les gestes et le mouvement du dispositif . . .	79
5.4.1	Expérimentation . . . . .	80
5.5	De l'utilisation du roulis . . . . .	83
5.5.1	Limite maximale de sélection et posture . . . . .	83
5.5.2	Approfondissement du cas à 7 éléments . . . . .	89
5.6	Conception de nouvelles techniques d'interaction . . . . .	90
5.7	Evaluation des techniques . . . . .	94
5.7.1	Expérimentation . . . . .	94
5.7.2	Résultats et discussion . . . . .	97
5.7.3	Mode débutant du menu . . . . .	100
5.7.4	Perspectives . . . . .	102
5.8	Conclusion . . . . .	103
6	BELLY GESTURES : INTERACTION GESTUELLE NON INSTRUMENTÉE	105
6.1	Introduction . . . . .	105
6.2	Belly gestures : exploration des gestes sur la surface de l'abdomen .	106
6.2.1	Les gestes sur la surface du corps . . . . .	106
6.2.2	Localisation des gestes sur le corps . . . . .	108
6.2.3	Propriétés de l'abdomen . . . . .	111
6.2.4	Propriétés de l'abdomen en tant qu'une surface interactive	113
6.2.5	Expérimentation : principe de l'expérience . . . . .	116
6.2.6	Équipement et procédure . . . . .	120
6.2.7	Résultats . . . . .	121
6.2.8	Analyse de l'exécution des gestes directionnels sur l'abdomen	130
6.2.9	Conclusion . . . . .	136
7	CONCLUSION ET PERSPECTIVES	139
7.1	Contributions . . . . .	139
7.1.1	Espaces de caractérisation . . . . .	140
7.1.2	Techniques d'interaction . . . . .	141
7.2	Perspectives . . . . .	142
7.2.1	Perspectives à court terme . . . . .	142
7.2.2	Perspectives à moyen terme . . . . .	142
7.2.3	Perspectives à long terme . . . . .	144
	BIBLIOGRAPHIE	145

---

## TABLE DES FIGURES

---

FIGURE 1.1	Logitech Revue : un clavier doté d'un trackpad permettent d'interagir avec Google TV . . . . .	2
FIGURE 2.1	Zenith Space Command 300, 1956. Deux boutons permettent de contrôler la chaîne et le volume. . . . .	6
FIGURE 2.2	Une télévision accompagnée d'une set-top box et de ses télécommandes. Depuis la TV, les téléspectateurs ont accès aux services fournis par Google (Google TV). . . . .	7
FIGURE 2.3	Le guide électronique des programmes de la Xbox One de Microsoft. . . . .	8
FIGURE 2.4	Catalogue de vidéo à la demande de Vudu pour la TV. . . .	9
FIGURE 2.5	Amigo TV : une communication vidéo entre amis durant la diffusion d'un programme. . . . .	10
FIGURE 2.6	Ecosystème d'applications de la Boxee box. . . . .	11
FIGURE 2.7	Interactions proxémiques : elles s'appuient sur la relation entre les utilisateurs et les dispositifs, les dispositifs entre eux et les objets non digitaux avec les utilisateurs et les dispositifs [Greenberg <i>et al.</i> , 2011]. . . . .	13
FIGURE 2.8	L'interface de SmartSkip [Drucker <i>et al.</i> , 2002]. . . . .	14
FIGURE 3.1	Télécommande SKY HD. Une télécommande traditionnelle.	19
FIGURE 3.2	TNT [Ingmarsson <i>et al.</i> , 2004]. Correspondance entre le clavier numérique et les caractères. . . . .	20
FIGURE 3.3	Télécommande Boxee équipée d'un clavier sur la face arrière.	21
FIGURE 3.4	Un exemple de télécommande doté d'un touchpad : la télécommande Panasonic Viera. . . . .	21
FIGURE 3.5	Un exemple de télécommande augmentée : les possibilités de la Wiimote. . . . .	22
FIGURE 3.6	Un exemple de télécommande vocale : la télécommande Ruwido R117. . . . .	24

FIGURE 3.7	Un exemple de télécommande en forme de cube : le lien entre la forme de la télécommande et la représentation visuelle est très étroit [Block <i>et al.</i> , 2004]. . . . .	25
FIGURE 3.8	Spinning Remote [Kimman <i>et al.</i> , 2011]. . . . .	26
FIGURE 3.9	Exploration de formes pour la manipulation de dispositifs (à gauche une sphère et au centre un cube ; à droite Smart-Its une plateforme de prototypage) [Block <i>et al.</i> , 2004]. . . . .	26
FIGURE 3.10	Peppermill [Villar <i>et al.</i> , 2010]. . . . .	27
FIGURE 3.11	<i>Rebo</i> : un geste sur le corps de la télécommande permet de changer de chaîne. [Kobayashi <i>et al.</i> , 2009]. . . . .	28
FIGURE 3.12	Kinect for Windows. . . . .	28
FIGURE 3.13	Utilisation de la tablette pendant l'activité télévisuelle (source : Institut de sondage Nielsen, Mai 2012). . . . .	30
FIGURE 3.14	Un pouf augmenté pour contrôler la télévision. [Schmidt <i>et al.</i> , 2004] . . . . .	31
FIGURE 3.15	Paper Remote : un guide de programme TV augmenté d'icônes cliquables avec un stylo numérique. [Berglund <i>et al.</i> , 2006] . . . . .	32
FIGURE 3.16	Un ballon capable de sentir les émotions. [Lee <i>et al.</i> , 2007] . . . . .	33
FIGURE 3.17	Une application de guide électronique des programmes et d'accès aux chaînes sur un dispositif mobile [Cruickshank <i>et al.</i> , 2007] . . . . .	35
FIGURE 3.18	G-Stalt : L'utilisateur manipule des vidéos à l'aide de gestes main-libre. [Zigelbaum <i>et al.</i> , 2010] . . . . .	36
FIGURE 3.19	Finger Count : Montrer ses mains et un nombre de doigt au système permet de sélectionner un élément dans un menu affiché à distance sur la TV [Bailly <i>et al.</i> , 2011] . . . . .	37
FIGURE 3.20	Twist&Tap : sélection de kanas en inclinant et appuyant sur les boutons de la télécommande [Aoki <i>et al.</i> , 2010]. . . . .	39
FIGURE 3.21	Un agent conversationnel pour la TV sous forme de peluche [Goto <i>et al.</i> , 2003]. . . . .	41
FIGURE 3.22	<i>Tactimote</i> : une cellule braille permet de transmettre de l'information par l'index. [Tahir <i>et al.</i> , 2008] . . . . .	42
FIGURE 4.1	Un geste symbolique avec les pouces levés vers le haut pour signaler que tout va bien. Photo de Andrew Beresford sur Flickr. . . . .	50

FIGURE 4.2	Une violoniste. Le geste d'excitation est illustré par la manipulation de l'archet sur les cordes, le geste de modulation par l'accord des doigts sur les cordes. Photo de Ben Clement sur Flickr. . . . .	51
FIGURE 4.3	Une proposition de caractérisation du geste dans le contexte du design industriel [Hummels et Stappers, 1998]. . . . .	52
FIGURE 4.4	Un geste manipulateur : l'utilisateur manipule un cube physique pour explorer l'interface affichée sur l'écran [Block <i>et al.</i> , 2004]. . . . .	54
FIGURE 4.5	Un utilisateur réalise un geste de défilement à la suite d'une démonstration de l'effet du geste [Wobbrock <i>et al.</i> , 2009]. . .	55
FIGURE 4.6	Fonction épistémique : l'utilisateur explore des textures tactiles virtuelles sur un objet réel [Bau et Poupyrev, 2012]. . .	62
FIGURE 4.7	Délimiteur dynamique : a - L'utilisateur pose les 5 doigts sur la surface pour calibration. b - Il retire les doigts non utilisés pour l'accord. c - Il réalise ensuite un mouvement directionnel [Lepinski <i>et al.</i> , 2010]. . . . .	65
FIGURE 4.8	Orientation du geste : le geste permet d'incliner le téléphone dans une direction pour sélectionner un élément du menu [Oakley et Park, 2009]. . . . .	67
FIGURE 4.9	Combinaison de gestes : L'utilisateur effectue un geste de rotation du téléphone dans un sens. Puis il effectue un geste de rotation dans l'autre sens. Le geste a une signature unique qui permet d'être reconnu sans la nécessité d'utiliser un délimiteur [Ruiz et Li, 2011]. . . . .	68
FIGURE 4.10	Combinaison séquentielle : le geste elliptique continu permet de contrôler le zoom sur une carte affichée sur un tableau interactif [Malacria <i>et al.</i> , 2010]. . . . .	69
FIGURE 4.11	Contexte social : étude de la préférence de réalisation des gestes dans un espace public [Rico et Brewster, 2010]. . . . .	71
FIGURE 5.1	Illustration des degrés de liberté du mouvement du dispositif.	74
FIGURE 5.2	Chameleon : A gauche, les translations du dispositif permet de se déplacer dans la grille de cellule et le temps. A droite, l'inclinaison du dispositif permet de sauter de cellule adjacente en cellule adjacente. . . . .	75

FIGURE 5.3	Marking Motion Menu. (a) L'utilisateur appuie contre l'écran tactile pour sélectionner le menu <i>contacts</i> . (b) Le menu <i>contacts</i> est affiché et l'utilisateur incline le dispositif pour sélectionner le sous-menu <i>messages</i> . (c) Le sous-menus <i>message</i> est affiché et l'utilisateur a repositionné le dispositif en position horizontale pour sélectionner la commande <i>inbox</i> . (d) Le contact avec l'écran est rompu pour actionner la commande. . . . .	75
FIGURE 5.4	<i>CubeKey</i> : les mouvements du dispositif dans l'espace en 3 dimensions correspondent directement au mouvement du curseur à l'intérieur du cube. . . . .	77
FIGURE 5.5	Gestext : disposition du clavier <i>tri-center</i> . Le curseur est initialement disposé au centre du clavier. Une simple rotation de roulis permet de sélectionner un caractère du bloc de gauche ou à droite. Une rotation de roulis de profondeur double sélectionne le centre du bloc de gauche ou de droite. . . . .	78
FIGURE 5.6	Expérience 1. A gauche, l'utilisateur incline le dispositif mobile pour atteindre la cible. A droite, l'utilisateur incline la télécommande pour atteindre la cible. . . . .	80
FIGURE 5.7	Exemple d'un geste de roulis. A gauche, un geste préparatoire pour augmenter l'amplitude disponible pour effectuer le geste. A droite, l'exécution du geste de roulis. La pression de la gâchette délimite le geste. . . . .	85
FIGURE 5.8	Stimulus pour la sélection d'une cible parmi sept en utilisant un geste de roulis. . . . .	86
FIGURE 5.9	Un exemple de mode novice de <i>Marking Menu</i> à deux niveaux avec 4 éléments affiché sur l'écran de la télévision. . .	91
FIGURE 5.10	Techniques retenues pour l'évaluation . . . . .	94
FIGURE 5.11	Stimulus graphique pour chaque modalité : à gauche bouton directionnel, au centre roulis, et à droite tangage et lacet. . . . .	95
FIGURE 5.12	Un utilisateur réalisant un geste de BD/Roulis pendant l'expérimentation. L'image du participant est en miroir car filmée par une caméra. . . . .	96
FIGURE 5.13	Temps moyen de sélection pour chaque combinaison . . . . .	97
FIGURE 5.14	Taux moyen de précision pour chaque combinaison . . . . .	98
FIGURE 5.15	Menu correspondant à une interaction T&L/Roulis. Le menu pour le roulis est un menu semi-circulaire de 5 éléments. . .	101

FIGURE 5.16	Représentation tabulaire des éléments. . . . .	101
FIGURE 6.1	<i>Shoesense</i> : un ensemble de gestes triangulaires [Bailly <i>et al.</i> , 2012]. . . . .	107
FIGURE 6.2	<i>Imaginary Phone</i> : l'utilisateur utilise la paume de sa main comme surface de substitution à son téléphone. Grace à la mémoire spatiale, l'utilisateur exécute une commande sur son téléphone en touchant la paume de sa main [Gustafson <i>et al.</i> , 2011]. . . . .	107
FIGURE 6.3	<i>Implanted user interfaces</i> : des capteurs sont intégrés sous une peau artificielle pour par exemple détecter une pression [Holz <i>et al.</i> , 2012]. . . . .	109
FIGURE 6.4	<i>Touché</i> : états détectés avec des gestes sur le corps et des capteurs portés sur les poignets. . . . .	110
FIGURE 6.5	Point de vue de la représentation de l'orientation spatiale . . . . .	117
FIGURE 6.6	Représentations internes de l'orientation spatiale identifiées chez les participants : exemple du chiffre "2" . . . . .	118
FIGURE 6.7	Exemple de stimuli utilisés pour le mode de présentation et le matériel présenté. Le point indique le point de départ. . . . .	118
FIGURE 6.8	Stimuli utilisés pour l'expérimentation : dans la partie supérieure, les stimuli pour le matériel directionnel ; dans la partie inférieure, les stimuli pour le matériel numéral . . . . .	119
FIGURE 6.9	Entropie moyenne (en bits) en fonction du mode de présentation et du matériel présenté. Les barres d'erreurs représentent un intervalle de confiance à 95%. . . . .	123
FIGURE 6.10	Taux des cas d'inversion par rapport à l'axe vertical. Les barres d'erreur représentent l'intervalle de confiance à 95%. . . . .	124
FIGURE 6.11	Taux des cas d'inversion par rapport à l'axe horizontal. Les barres d'erreur représentent l'intervalle de confiance à 95%. . . . .	126
FIGURE 6.12	Temps moyen de réaction par participant. Les barres d'erreur représentent l'intervalle de confiance à 95%. . . . .	127
FIGURE 6.13	Temps d'exécution médians en secondes. . . . .	132
FIGURE 6.14	Amplitude médiane du geste en millimètres. . . . .	132
FIGURE 6.15	Vitesse médiane du geste (en mètres par seconde). . . . .	133
FIGURE 6.16	Linéarité moyenne du geste : valeur moyenne des maxima du coefficient de détermination $r^2$ de la régression linéaire des tracés pivotés sur $360^\circ$ . . . . .	134

FIGURE 6.17	Erreur directionnelle moyenne du geste (en degrés). Les barres d'erreur représentent l'intervalle de confiance à 95%. . . . .	134
FIGURE 6.18	Variabilité directionnelle du geste (en degrés) : moyenne des valeurs absolues de l'erreur directionnelle . . . . .	135

---

## LISTE DES TABLEAUX

---

TABLE 3.1	Catégorisation des travaux par type de télécommande et dimension interactionnelle . . . . .	47
TABLE 4.1	Espace de caractérisation du geste de la main pour la télévision interactive . . . . .	59
TABLE 5.1	Variation angulaire moyenne pour chaque élément par taille de groupe . . . . .	88
TABLE 5.2	Temps moyen de sélection (en secondes) et amplitude angulaire moyenne pour chaque élément (en degrés). . . . .	90
TABLE 5.3	Considération des techniques de l'espace de conception pour évaluation : les techniques en fond gris ne sont pas retenues pour évaluation. . . . .	93
TABLE 5.4	Préférences subjectives moyennes sur une échelle de 1 à 7. . . . .	99
TABLE 6.1	Fréquence moyenne des transformations pour chaque participant et pour chaque condition expérimentale . . . . .	125

---

## INTRODUCTION

---

### 1.1 CONTEXTE GÉNÉRAL

Depuis son apparition dans les années 40, la télévision n'a cessé de se populariser. Récemment, l'évolution technologique des décodeurs (set-top boxes), maintenant capables d'accéder à Internet, a ouvert de nouvelles possibilités pour améliorer l'expérience télévisuelle. Grâce à cette convergence entre la télévision et Internet, les fournisseurs proposent de plus en plus de services dont l'interactivité ne cesse de croître.

Contrairement à la situation où il utilise un ordinateur avec un clavier et une souris, le téléspectateur interagit à distance de la télévision à l'aide d'une télécommande et souhaite accéder facilement aux contenus qui l'intéresse, confortablement assis dans son canapé. La télécommande traditionnelle dont l'expressivité est limitée ne permet pas aisément de naviguer dans les menus hiérarchiques complexes ou les larges catalogues de contenu mise à disposition des téléspectateurs, ou de saisir de l'information. De nombreux constructeurs<sup>1</sup> se sont intéressés à cette problématique, sans vraiment tenir compte des particularités du contexte de l'interaction avec la télévision, comme en témoigne l'échec de Google TV<sup>2</sup> par exemple (Figure 1.1).

Les différences entre l'usage de la télévision et de l'ordinateur de bureau impliquent de reconcevoir la manière d'interagir avec la télévision tout comme cela a été le cas pour les dispositifs mobiles. La conception de techniques d'interaction adaptées aux spécificités matérielles et à l'usage des smartphones ont accéléré l'adoption en masse de ces dispositifs et des applications qu'ils offrent. Il pourrait en être de même pour la télévision interactive.

Cette thèse, réalisée avec le soutien du projet *QUAERO*<sup>3</sup>, s'intéresse aux techniques d'interaction pour la télévision interactive. Plus particulièrement,

---

1. [http://www.maclife.com/article/analysis/3\\_worthy\\_alternatives\\_apple\\_tv](http://www.maclife.com/article/analysis/3_worthy_alternatives_apple_tv)

2. <http://bits.blogs.nytimes.com/2012/05/07/google-tries-again-with-google-tv-2/>

3. <http://www.quaero.org/>



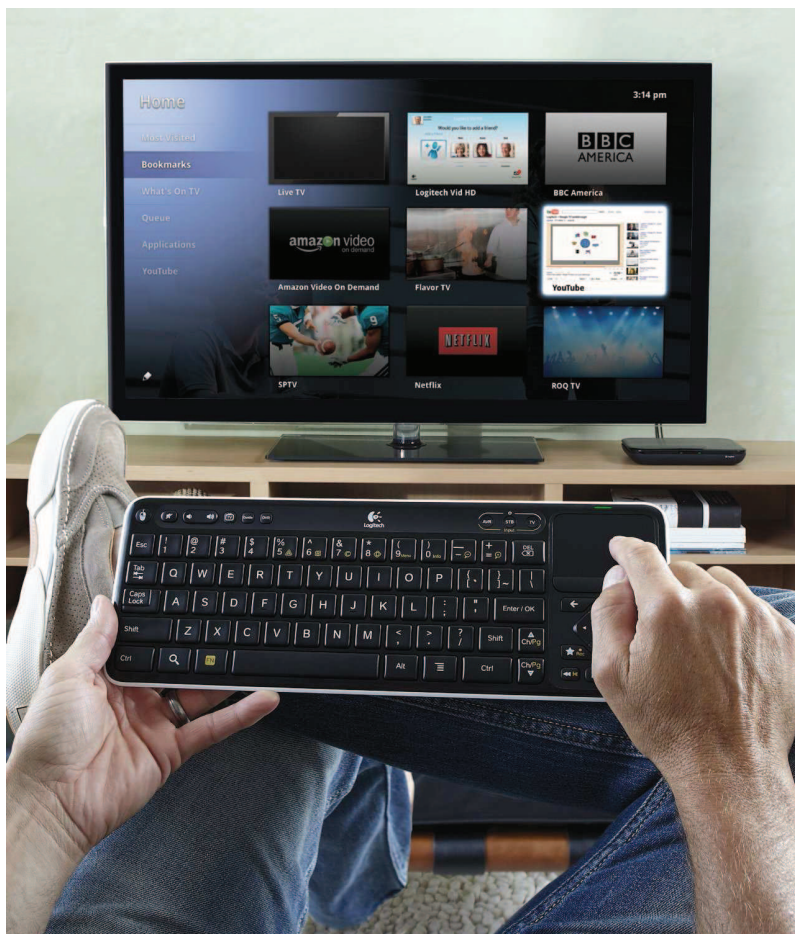


FIGURE 1.1: Logitech Revue : un clavier doté d'un trackpad permettent d'interagir avec Google TV

elle présente des travaux qui tentent d'améliorer les techniques d'interaction pour la télévision et la télécommande traditionnelle en tenant compte du contexte d'usage dans lequel les téléspectateurs regardent la télévision.

## 1.2 ORGANISATION DU MÉMOIRE

En préambule à cette thèse, nous présentons dans le chapitre 2 un aperçu de l'évolution de la télévision au fil des dernières décennies pour introduire la notion de télévision interactive. En particulier, ce chapitre vise à donner une vue d'ensemble des nombreux services aujourd'hui accessibles par ce biais. Il aborde également les différentes caractéristiques qui font sa singularité et la

différencient d'une activité productive.

Dans le chapitre 3, nous revenons d'abord sur la télécommande traditionnelle afin d'en décrire l'évolution et en constater les limites actuelles. Puis, nous présentons un espace de classification de l'ensemble des travaux visant à l'améliorer pour simplifier l'accès et le contrôle des services ou enrichir l'expérience des téléspectateurs. Cet espace nous conduira à nous intéresser à l'interaction gestuelle.

Le chapitre 4 est consacré à la caractérisation de l'interaction gestuelle avec la main. Après une revue des différentes taxonomies existantes sur les gestes, nous proposons un espace de classification qui analyse les gestes selon cinq dimensions : le type d'interaction (geste sur surface ou mains libres, avec ou sans instrument), la fonction du geste, l'anatomie du geste, la combinatoire du geste et le contexte dans lequel il est réalisé. Cet espace met en exergue différentes propriétés utilisables pour augmenter l'expressivité du geste, par exemple la possibilité de les combiner entre eux et ainsi augmenter la bande passante interactionnelle. Nous présentons la conception de techniques d'interaction qui s'appuient sur ces propriétés dans les chapitres suivants.

Le chapitre 5 s'intéresse à l'enrichissement de l'interaction avec une télécommande par des gestes en l'air pour exécuter des commandes fréquentes ou accéder à ses contenus favoris rapidement et facilement. Après une revue de la littérature sur l'utilisation des mouvements d'un dispositif pour l'interaction gestuelle, nous explorons les dimensions du geste qui ne sont pas utilisées par les téléspectateurs lorsqu'ils manipulent une télécommande pour concevoir de nouvelles techniques d'interaction (en combinant boutons de la télécommande et gestes). Ces techniques font l'objet d'une évaluation comparative.

Le chapitre 6 présente *Belly Gestures* une technique d'interaction gestuelle qui s'abstrait de tout dispositif physique pour exécuter rapidement des commandes. Cette technique repose essentiellement sur des gestes de la main sur le ventre et la proprioception. Dans ce chapitre, nous proposons d'explorer les caractéristiques de la surface du ventre et d'étudier la représentation mentale spatiale des utilisateurs lorsqu'ils interagissent pour différents types de gestes. Puis nous consacrons une partie de ce chapitre à l'étude de l'exécution motrice de ces gestes.

Enfin, le dernier chapitre conclut ce mémoire et présente quelques perspectives de recherche.

---

## CONTEXTE

---

### 2.1 DE LA TÉLÉVISION À LA TÉLÉVISION INTERACTIVE

Depuis son introduction dans les ménages dans les années 40, la télévision (TV) est devenue une source de communication d'information et de divertissement très populaire. Pour preuve, en 2012, le Conseil Supérieur de l'Audiovisuel reportait que 98,3% des foyers français étaient équipés d'au moins un poste de TV et chaque français regardait cet écran durant 3 heures et 50 minutes en moyenne par jour <sup>1</sup>. Au cours des années, ce système audio visuel a évolué pour fournir des contenus multimédia (programmes TV, vidéos, applications dédiées...) toujours plus riches que les téléspectateurs consomment avec une attention variable, individuellement, en famille ou entre amis. Plus récemment, la possibilité de connecter sa TV à Internet a ouvert de nouvelles opportunités telles que la personnalisation de la diffusion des programmes, l'accès à des médias destinés à d'autres plate-formes de diffusion, la participation active des téléspectateurs avec le contenu de diffusion ou encore le partage délocalisé de l'expérience télévisuelle.

En parallèle, depuis son apparition dans les années 1950 <sup>2</sup> (Figure 2.1), la télécommande sans fil, qui permet de contrôler la TV à distance, a évolué à un rythme modéré. Avec l'apparition continue de nouvelles fonctionnalités sur la TV, le nombre de boutons habillant la télécommande a vu son nombre augmenter de même que l'attention visuelle qui lui est dédiée, complexifiant ainsi son utilisation. Ce phénomène a aussi entraîné l'apparition de menus linéaires hiérarchiques sur les télévisions dont, contrairement aux interfaces d'ordinateur de bureau équipé d'un clavier et d'une souris, la navigation à la télécommande traditionnelle est fastidieuse. Par ailleurs, la commercialisation de matériels de décodage de fréquence (set-top box), d'enregistrement de programmes (magnétoscopes) ou de lecture de médias physiques (VHS, DVD, Bluray...) possédant chacun leur protocole de communication propriétaire a

---

1. Chiffre clés de l'audiovisuel français, 1er semestre 2013. <http://bit.ly/chiffresclésaudiovisuel2013>

2. <http://www.zenith.com/remote-background>



FIGURE 2.1: Zenith Space Command 300, 1956. Deux boutons permettent de contrôler la chaîne et le volume.

contribué à la multiplication des télécommandes avec différentes configurations dans le salon.

Ces évolutions successives, à la fois de la TV et de la télécommande, ont étendu les possibilités offertes aux téléspectateurs. Cependant, la navigation au sein de l'écosystème et l'utilisation de ces services avec l'expressivité limitante de la télécommande a complexifié l'usage de la TV au détriment de la simplicité et de l'efficacité. Dans ce chapitre, nous présentons une définition de la TV interactive et certaines caractéristiques de l'activité des téléspectateurs.

## 2.2 LA TÉLÉVISION INTERACTIVE

[Chorianopoulos \[2008\]](#) définit la TV interactive comme une classe de services qui pourraient transformer l'activité avec la TV dans le salon (Figure 2.2). Elle est également un moyen permettant aux téléspectateurs de dépasser l'expérience passive de regarder la TV et leur donner la possibilité de faire des choix et d'agir [\[Gawlinski, 2003\]](#). Plus précisément, la TV interactive est une forme de TV qui se base sur des interactions avec le média sous forme de choix, de décisions et de communication. De cette manière, il devient possible de contrôler ce qu'on regarde, quand, comment, de participer activement aux programmes diffusés ou de partager du contenu généré par les utilisateurs avec les téléspectateurs [\[Jensen, 2008\]](#).



FIGURE 2.2: Une télévision accompagnée d'une set-top box et de ses télécommandes. Depuis la TV, les téléspectateurs ont accès aux services fournis par Google (Google TV).

Aujourd'hui, la TV est généralement couplée à une set-top box mise à disposition par un fournisseur de contenu qui permet d'accéder à une pléthore de chaînes et divers services interactifs. Ces services peuvent appartenir à différentes catégories décrites ci-après : les guides de programmes électroniques, les programmes augmentés affichant des informations par dessus les programmes, la vidéo à la demande, la TV personnalisée, les services de t-commerce, les jeux, les services de TV sociales et les services d'information [Jensen, 2005].

Le nombre de chaînes proposés aux téléspectateurs ne cesse de croître. En France, le Conseil Supérieur de l'Audiovisuel recensait près de 216 chaînes accessibles par divers réseaux de diffusion<sup>3</sup> en plus des 32 chaînes de la télévision numérique terrestre<sup>4</sup>. Cette quantité de chaînes et les programmes qu'elles diffusent rendent la navigation et la sélection difficile à tel point que le zapping de chaîne en chaîne et l'acquisition d'information devient impraticable [Black *et al.*, 1994].

Les guides de programmes électroniques (EPG) [Berglund *et al.*, 2006; Jensen, 2008] permettent d'explorer la liste des programmes disponibles (Figure 2.3). Ils peuvent comprendre un moteur de recherche textuel interac-

3. [http://bit.ly/chaines\\_autres\\_reseaux](http://bit.ly/chaines_autres_reseaux)

4. [http://bit.ly/chaines\\_tnt](http://bit.ly/chaines_tnt)

tif, des fonctionnalités de rappel de diffusion ou d'enregistrement de programmes, ou encore des outils de personnalisation. Ces guides tendent à devenir de plus en plus conséquents avec l'augmentation constante du nombre de chaînes disponibles. Ils proposent généralement une navigation et une saisie de texte basée sur la télécommande qui s'avère inefficace pour ces tâches [Berglund *et al.*, 2006]. La navigation à travers ces guides requiert des moyens d'accéder plus rapidement à l'information recherchée.

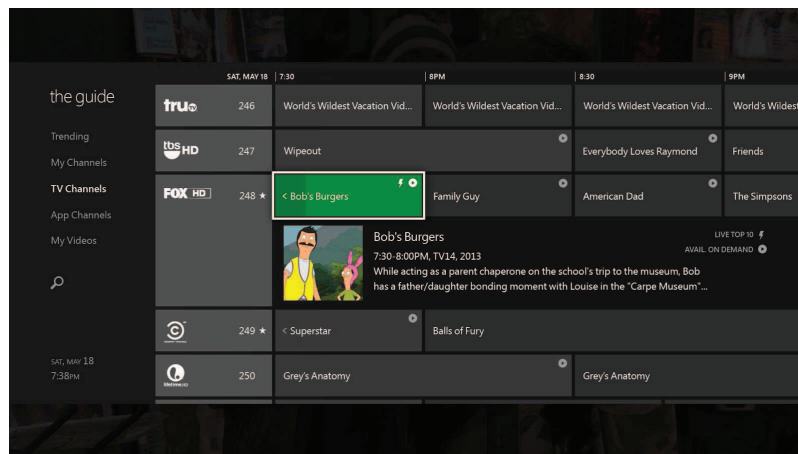


FIGURE 2.3: Le guide électronique des programmes de la Xbox One de Microsoft.

Les programmes augmentés donnent accès à des informations supplémentaires lors du visionnage d'un programme. Ces informations peuvent être en relation avec le programme, comme l'affichage du résumé ou un flux de commentaires des téléspectateurs en direct, ou indépendantes comme l'affichage des dernières actualités ou de la météo. Ces informations peuvent occulter le contenu diffusé. Avec la popularisation de l'usage des tablettes tactiles en complément de l'écran de télévision, l'affichage de ces informations tend à être déporté sur ce support, mais modifie la distribution de l'attention visuelle pendant l'activité télévisuelle [Holmes *et al.*, 2012]. Ces dispositifs accaparent l'attention visuelle des téléspectateurs qui se désengagent du programme diffusé sur l'écran de la télévision.

La vidéo à la demande donne la possibilité aux téléspectateurs de choisir du contenu vidéo et le moment auquel il souhaite le regarder. Deux types de service leurs sont offerts : la location ou l'achat de contenus vidéo de différents distributeurs et la TV de rattrapage (catch up TV) permettant de revoir une émission déjà diffusée durant une période limitée. Ils sont généralement matérialisés par des catalogues virtuels conséquents dans lequel le téléspectateur navigue à l'aide de la télécommande pour faire son choix (Figure 2.4). De



même que pour le guide des programmes électroniques, la navigation dans ces immenses catalogues par le biais des boutons de la télécommandes est fastidieux tout comme la saisie de texte dans les champs de recherche. L'exploration de ces catalogues au choix abondant est également un défi car elle doit permettre aux téléspectateurs de découvrir du contenu susceptible de l'intéresser.

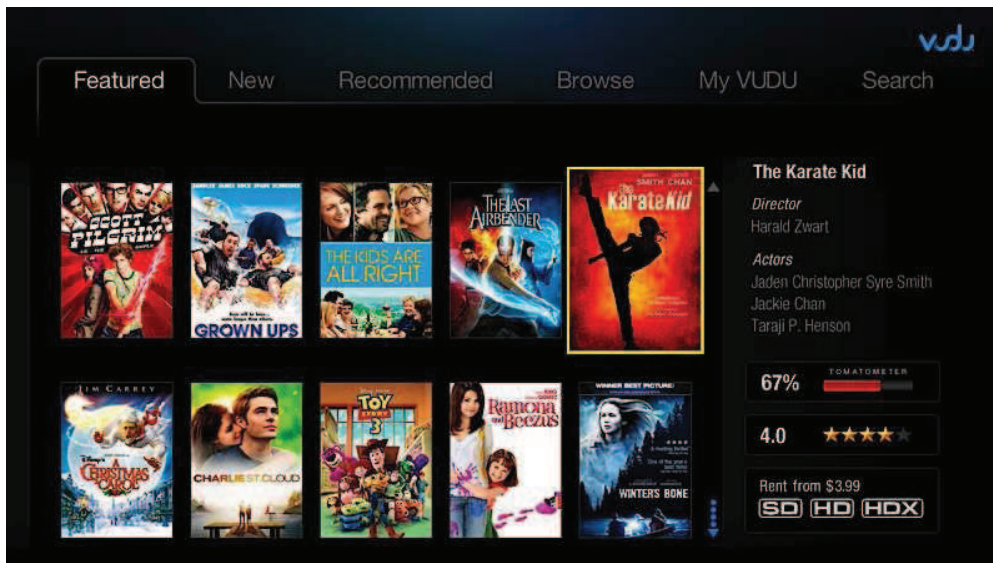


FIGURE 2.4: Catalogue de vidéo à la demande de Vudu pour la TV.

La TV personnalisée fournit un contrôle plus précis sur le contenu diffusé. Elle permet en particulier d'enregistrer un programme pour un visionnage différé d'un court instant, suite à une interruption, ou d'une durée indéterminée. Ces enregistrements permettent aussi de rejouer un programme et de se déplacer temporellement dans le flux. Ceci permet, par exemple, d'éviter les publicités ou de revoir un moment important d'un programme [Drucker *et al.*, 2002]. Cette approche offre aussi un contrôle plus fin sur le flux de diffusion d'un programme. Par exemple, pendant la retransmission d'un événement sportif, les téléspectateurs peuvent rejouer le ralenti d'une action clé sous plusieurs angles différents [Olsen *et al.*, 2010].

La TV sociale est définie comme une activité qui met en avant l'interaction synchronisée entre les occupants des salons interconnectés entre eux (par le réseau du diffuseur des contenus) [Svensson et Sokoler, 2008]. Le système AmigoTV<sup>5</sup> d'Alcatel illustre les possibilités offertes par cette approche.

5. <http://bit.ly/amigotv>



Entre autres, ce système permet de notifier la présence d'amis connectés sur le réseau et communiquer en temps réel entre téléspectateurs pendant la diffusion d'un programme par le biais de conversations audio ou vidéo ou de messages iconographiques visuels ou sonores (Figure 2.5). Des services supplémentaires peuvent être envisagés comme les services de recommandation de programmes ou encore les canaux de réseaux sociaux déjà populaires sur internet.

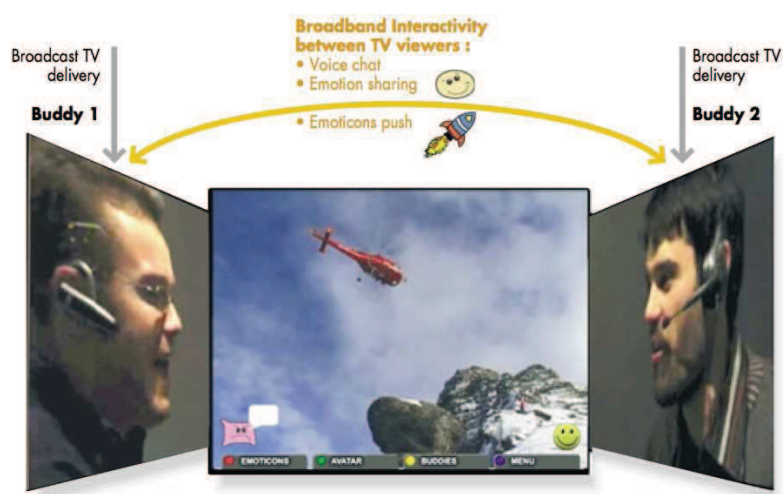


FIGURE 2.5: Amigo TV : une communication vidéo entre amis durant la diffusion d'un programme.

Les services de t-commerce (commerce télévisé) permettent de commander des produits directement de la TV pendant la diffusion d'une émission de téléachat. D'autres services indépendants des programmes diffusés comme les services bancaires sont également accessibles. Il est alors possible de consulter l'état de ses finances ou d'effectuer des transactions en utilisant sa télécommande.

Les jeux permettent d'engager davantage les téléspectateurs avec les programmes pendant leur diffusion. Ces jeux peuvent prendre la forme de quizz en relation directe avec le contenu de l'émission. Les fournisseurs de contenu offrent aussi la possibilité de télécharger des jeux pour jouer tout seul ou en groupe avec les autres joueurs présents sur le réseau du fournisseur.

Enfin, divers services Internet peuvent être rendus accessibles directement sur le téléviseur. A l'image des smartphones, les fournisseurs offrent des écosystèmes d'applications pour accéder à ses messages électroniques, aux actualités sélectionnées, aux critiques de films, tout autre contenu qui pourrait

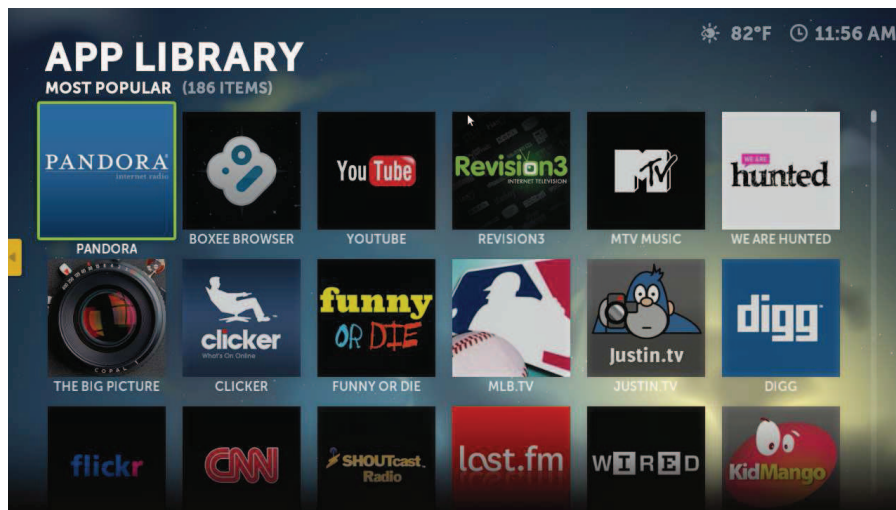


FIGURE 2.6: Ecosystème d'applications de la Boxee box.

intéresser le téléspectateur (Figure 2.6).

Cette prolifération et cette complexification des services interactifs modifie le concept vieillissant de la TV. Le rôle du téléspectateur devient moins passif et son engagement vers la TV plus important. Il a maintenant l'opportunité de choisir parmi un large éventail de contenus et d'interagir avec les programmes et les autres téléspectateurs. De nouvelles formes d'interaction sont alors nécessaires pour faciliter la saisie de texte, la navigation ou l'accès au contenu. Nous y reviendrons dans le chapitre suivant.

## 2.3 UNE ACTIVITÉ AVEC DIFFÉRENTES CARACTÉRISTIQUES

Comme nous l'avons vu précédemment, la TV interactive engage de plus en plus les téléspectateurs en les rendant plus actifs avec le contenu. Elle intervient dans plusieurs contextes. Tout d'abord on peut regarder la TV seul ou à plusieurs. Ensuite, l'attention accordée à la TV est partagée entre plusieurs activités concurrente et donc variable. Enfin, la TV est une activité associée au divertissement et diffère d'un contexte de travail. Nous abordons ces spécificités ci-après.

### 2.3.1 Une activité sociale

L'activité télévisuelle est souvent partagée entre amis ou membres d'une même famille. L'utilisation de services interactifs peut alors s'avérer problé-

matique si l'accès à un service accapare une partie ou la totalité de la TV, ceci au détriment des autres utilisateurs. La prise de contrôle du dispositif peut également être source de tension, *Meerbeek et al.* [2010] suggérant que cette question du contrôle pèse sur la satisfaction des utilisateurs. De plus, des problèmes de vie privée peuvent surgir si l'échange d'informations entre un utilisateur et le service interactif est visible par les autres personnes [*Cesar et al.*, 2009]. L'utilisation d'un dispositif auxiliaire [*Cesar et al.*, 2009; *Cruickshank et al.*, 2007; *Tsekleves et al.*, 2007] comme un téléphone portable ou une tablette peut alors constituer une solution en offrant un espace d'interaction privé opaque aux autres utilisateurs ceci évitant à la fois de les gêner et d'éventuelles indiscretions [*Cruickshank et al.*, 2007; *Tsekleves et al.*, 2007].

Certains travaux de recherche prennent en considération le degré d'engagement des utilisateurs et l'activité sociale dans la conception des systèmes [*Gueddana et Roussel*, 2009; *Ballendat et al.*, 2010]. Ils pourraient être adaptés dans le contexte de la télévision. Pêle-Mêle est un système de communication vidéo multi-sites qui utilise la détection de présence, de mouvement et de visage pour classifier en permanence le niveau d'activité des utilisateurs [*Gueddana et Roussel*, 2009]. Ce niveau d'activité est mesuré selon trois niveaux (absent, disponible, engagé) qui impactent le contenu : les images des lieux de niveau "engagé" apparaissent au centre de l'écran, les autres étant relégués vers la périphérie. Ils impactent aussi la fréquence de mise à jour des images : si le niveau d'engagement est "absent", l'image est figée et dégradée. La transition et combinaison avec d'autres services est aussi influencée : des animations permettent de faciliter la perception et la compréhension des transitions entre les différents niveaux d'engagement.

*Ballendat et al.* [2010] ont introduit et appliqué la notion d'interaction proxémique au contexte de la télévision : selon le lieu, la distance, l'orientation, le mouvement des utilisateurs par rapport au dispositif, et selon l'identité de ce dispositif, le système offre différentes options (Figure 2.7). Par exemple, si l'utilisateur s'approche du téléviseur, le système adapte la visualisation afin de faciliter l'exploration des vidéos disponibles. S'il s'éloigne pour s'asseoir dans son canapé, la vidéo sélectionnée passe en plein écran et sa diffusion commence. Inversement, la vidéo s'arrête si l'utilisateur reçoit un appel téléphonique ou ouvre un magazine. Si un autre participant s'approche de l'écran, des informations sont affichées à son attention mais l'écran est partagé de manière à ne pas interrompre le flux intéressant le participant assis. Le système met en pause la vidéo si les participants rentrent en discussion et se met en veille s'ils quittent la pièce. D'autres fonctionnalités permettent aussi de détecter si l'utilisateur pointe l'écran avec un objet (pour interagir

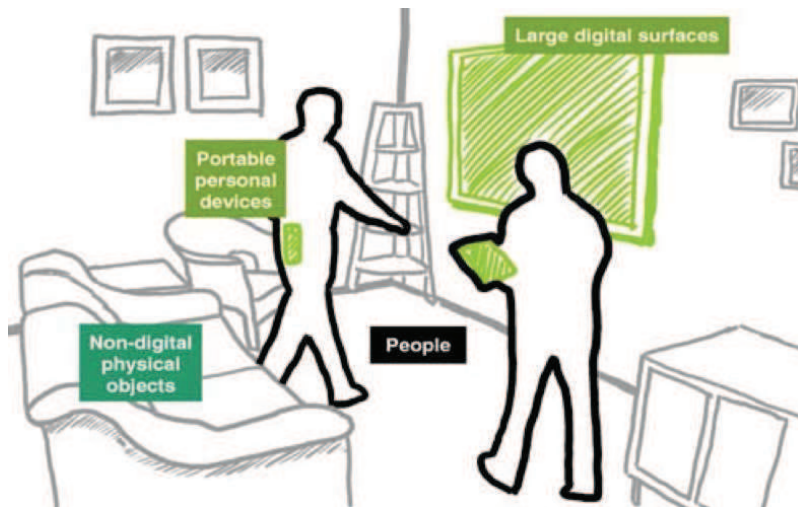


FIGURE 2.7: Interactions proxémiques : elles s'appuient sur la relation entre les utilisateurs et les dispositifs, les dispositifs entre eux et les objets non digitaux avec les utilisateurs et les dispositifs [Greenberg *et al.*, 2011].

avec), s'il approche un téléphone mobile près de l'écran (pour transférer des contenus multimédia), etc. D'autres exemples de tels systèmes figurent dans dans [Greenberg *et al.*, 2011].

### 2.3.2 Une activité à différents degrés d'engagement

Le téléspectateur peut naturellement mener d'autres activités en parallèle. Berglund *et al.* [2006] identifient quatre types d'usage de la TV en fonction du niveau d'engagement : activité d'arrière plan, d'accompagnement, partielle ou principale. L'activité d'arrière plan traduit le fait que la TV est allumée, mais que personne ne la regarde. Regarder la TV est une activité d'accompagnement lorsque l'utilisateur est engagé dans une autre tâche et tout à coup son attention est focalisée sur la TV. C'est le cas, par exemple, lors de la retransmission d'un match de football quand une équipe marque un but. Elle est une activité partielle quand l'utilisateur partage cette tâche alternativement avec une autre. Repasser son linge en regardant la TV en est un exemple. Enfin, elle est l'activité principale quand toute l'attention du téléspectateur est portée sur la TV.

Gueddana et Roussel [2009] définissent le niveau d'engagement comme « la limite pour laquelle une personne est prête à s'exposer et à être sollicitée par d'autres ». Lorsque certaines modalités sensorielles sont requises dans une

autre tâche, elles sont plus ou moins indisponibles pour la TV. Ce facteur de niveau d'engagement est un facteur critique pour la conception des applications pour la TV [Chorianopoulos, 2008]. Par exemple, dans [Chorianopoulos, 2008] une application de diffusion de clips musicaux, le flux de la diffusion vidéo est continu, sauf si le bouton « chanson suivante » est pressé. Cette application transforme de fait le téléviseur en une sorte de radio, permettant à l'utilisateur de se consacrer à une autre activité.

### 2.3.3 Une activité de divertissement

L'objet de l'activité télévisuelle est de se divertir, de s'informer et de se détendre [Bernhaupt *et al.*, 2008; Chorianopoulos et Spinellis, 2006], confortablement installé dans un canapé plutôt qu'assis devant un écran d'ordinateur [Bernhaupt *et al.*, 2008; Tsekleves *et al.*, 2011]. Bernhaupt *et al.* [2008] rapportent que les téléspectateurs ne sont pas disposés à utiliser clavier et souris, et qu'ils sont même réticents à l'idée d'une interface graphique exploitant la métaphore du bureau. Les modèles qui ont fait le succès de l'ordinateur personnel classique semblent donc peu adaptés au contexte de la TV interactive.



FIGURE 2.8: L'interface de SmartSkip [Drucker *et al.*, 2002].

Il y a lieu dès lors de s'interroger sur la pertinence, par rapport à notre sujet, des méthodes d'évaluation traditionnelles en IHM qui reposent sur des mesures de performance – e.g., un temps d'exécution de tâche que l'on demande de minimiser [Chorianopoulos et Spinellis, 2006; Drucker *et al.*, 2002].

Les critères d'évaluation des techniques d'interaction pour les applications productives ne suffisent pas pour celles de la TV (interactive) [Chorianopoulos et Spinellis, 2006]. Drucker *et al.* [2002] rapportent que bien que les performances des utilisateurs avec leur système de navigation vidéo *SmartSkip* ne soient pas les meilleures, ceux-ci jugent leur système plus amusant, moins frustrant et aussi plus facile à apprendre et à utiliser que les deux autres systèmes comparés (Figure 2.8). Les auteurs suggèrent que les préférences des utilisateurs et leurs performances dépendent du niveau d'engagement face à la TV. Chorianopoulos et Spinellis [2006] proposent une méthodologie d'évaluation basée sur les émotions et prenant en compte les caractéristiques particulières de la TV, des téléspectateurs et du contexte d'usage. Outre l'examen traditionnel de l'efficacité des interfaces, ils proposent de réutiliser des méthodes d'évaluation déjà éprouvées dans d'autres domaines, comme l'étude des média et de la communication de masse [Knobloch et Zillmann, 2002].

## 2.4 CONCLUSION

Dans ce chapitre nous avons défini la TV interactive et son contexte d'activité. La TV passive des années 40 a laissé place à une source de divertissement où l'engagement avec les programmes et les autres téléspectateurs, et le contrôle du contenu sont devenus de plus en plus présents. Les possibilités d'interaction avec les programmes, les catalogues de contenu et les autres téléspectateurs ont enrichi l'expérience télévisuelle. Les nouvelles tâches relatives à ces nouveaux usages demandent de revoir le concept de la télécommande traditionnelle qui, en l'état, n'est plus adapté à une telle variété de possibilités interactionnelles. Ce travail a déjà été amorcé par la communauté en interaction homme-machine. Dans le chapitre suivant, nous montrons la limitation évidente de la télécommande traditionnelle pour ces nouveaux usages et explorons les travaux qui transforment la télécommande familière pour en simplifier leur usage.





---

## ESPACE DE CARACTÉRISATION DES TÉLÉCOMMANDES

---

Dans le chapitre précédent, nous avons précisé la notion de télévision (TV) interactive. Nous avons notamment vu qu'elle vise à donner accès à de nombreux services interactifs. La télécommande est un dispositif d'entrée permettant de contrôler ces divers services et de manipuler les données multimédia mises à disposition de l'utilisateur. Il est frappant de constater qu'elle a beaucoup moins évolué que les services offerts à l'utilisateur [Cooper, 2008]. De fait, la télécommande traditionnelle se révèle insuffisante pour naviguer, contrôler et profiter de ces nouveaux services aisément. La littérature fait état de nombreuses tentatives pour améliorer la télécommande.

Pour classifier ces travaux, Cesar *et al.* [2008] considèrent trois catégories : les télécommandes traditionnelles augmentées, l'utilisation opportuniste des terminaux mobiles personnels (smartphones, tablettes tactiles) et la réutilisation des objets du quotidien. L'augmentation de la télécommande traditionnelle consiste à équiper le dispositif de nouveaux capteurs afin de le pourvoir de nouvelles capacités interactionnelles. Cette approche permet par exemple d'introduire la modalité gestuelle tout en conservant la familiarité et les usages déjà ancrés de la télécommande classique. Par ailleurs, la popularisation des terminaux mobiles, que ce soit le smartphone ou la tablette tactile, permet aux utilisateurs de rester continuellement connectés à Internet. Dans le salon, ces terminaux peuvent offrir une expérience télévisuelle plus riche en donnant accès à du contenu additionnel. Elles peuvent de plus servir de support à des techniques d'interaction avancées. Enfin, de nombreux artefacts en relation ou dans l'environnement de la TV sont manipulés quotidiennement. Par exemple, les programmes et magazines papier de TV accompagnent couramment la planification des programmes à regarder. Les travaux qui tirent avantage de cette familiarité et de ces routines pour faciliter l'utilisation des services matérialisent une troisième catégorie : la réutilisation des objets du quotidien qui sont augmentés.

Cette première trichotomie catégorise différentes stratégies pour faciliter l'accès et les usages des services interactifs. En nous inspirant des travaux



de Cesar *et al.*, nous proposons d'étendre cette classification en considérant deux types de télécommandes supplémentaires. La littérature présente un nombre conséquent de travaux qui tentent de reconcevoir la télécommande en tenant compte des aptitudes des téléspectateurs à manipuler des objets physiques. Ces travaux correspondent à une catégorie que l'on peut appeler télécommande tangible. A la différence des télécommandes augmentées qui sont fortement basées sur l'apparence et la forme de la télécommande traditionnelle, les télécommandes tangibles s'appuient sur des formes, une apparence et des métaphores d'interaction proposant un lien étroit avec la représentation des données à laquelle elles sont associées. Enfin, l'usage d'une télécommande physique comporte certains désavantages concernant sa disponibilité, son entretien ou sa solidité. Pour contrer ces désavantages, un pan de la recherche propose d'investiguer des techniques d'interaction s'abstenant de toute manipulation d'un dispositif physique. Nous appelons cette catégorie les télécommandes dématérialisées.

Dans ce chapitre, nous présentons d'abord la télécommande traditionnelle, son évolution jusqu'à aujourd'hui et ses limitations actuelles. Puis, nous abordons les différents types de télécommande proposés afin d'améliorer la télécommande traditionnelle, introduits précédemment, en distinguant les télécommandes dédiées à la TV et celles servant également à d'autres activités. Enfin, nous explorons les différentes modalités d'entrée et de sortie impliquées dans la conception de ces prototypes. Un récapitulatif de cette classification est présenté au tableau 3.1.

### 3.1 LA TÉLÉCOMMANDE TRADITIONNELLE

La télécommande traditionnelle (Figure 3.1) est composée de touches dédiées à la navigation, aux fonctionnalités courantes, ou encore à la saisie de données. Elle tend à être surchargée de boutons à cause de l'augmentation et de la richesse des services interactifs [Cooper, 2008; Pirker *et al.*, 2010]. En plus d'accaparer l'attention visuelle normalement dédiée à la TV, sa difficulté d'utilisation [Bernhaupt *et al.*, 2008] est un frein à l'adoption des nouvelles fonctionnalités par les utilisateurs [Pirker *et al.*, 2010].

Pour pallier à la surcharge de boutons, une solution offerte par les télécommandes classiques consiste à multiplier les modes, un par appareil physique, et les menus dans lesquels il faut naviguer au moyen de nombreux appuis successifs sur des touches directionnelles [Berglund et Johansson, 2004; Chorianopoulos et Spinellis, 2006]. Contrairement au cas de l'interaction classique

avec une souris et un clavier, la navigation dans des menus hiérarchiques tend à être plus laborieuse avec une télécommande [Chorianopoulos et Spinellis, 2006]. Mirlacher *et al.* [2010] montrent que six touches sur la télécommande (4 touches directionnelles, plus 1 pour valider et 1 pour annuler) est un bon compromis entre simplicité et efficacité pour naviguer dans les menus hiérarchiques. L'offre de services interactifs toujours croissante impose de trouver des techniques pour accéder au contenu tout en limitant le nombre de boutons et la complexité des menus.



FIGURE 3.1: Télécommande SKY HD. Une télécommande traditionnelle.

Certains attributs des boutons, comme leur couleur ou leur forme, peuvent être mis à profit pour réaliser des actions soit globales soit relatives au mode ou à la sélection courante [Cooper, 2008]. Une telle ressource est toutefois d'une portée limitée et n'excède pas quatre couleurs en pratique (Figure 3.1), à plus forte raison si l'on garde à l'esprit que la TV se regarde généralement à faible niveau de luminosité ambiante. Les distinctions de forme évitent ce problème et ont l'avantage d'être reconnaissables par un simple toucher, ceci permettant d'activer des fonctionnalités sans avoir à déplacer le regard de la TV vers la télécommande. Enfin, les touches numériques permettent d'accéder directement aux chaînes (si tant est que l'utilisateur se remémore le nombre qui leur est associé) ou parfois à naviguer dans les menus hiérarchiques (chaque élément étant alors associé à un chiffre).

L'entrée de texte constitue un autre problème car le facteur de forme interdit généralement de disposer d'un clavier classique [Marshall *et al.*, 2001]. De même que pour les téléphones mobiles, la technique du *Multitap* peut permettre d'entrer du texte si la télécommande dispose d'un clavier numérique [Cooper, 2008]. Une alternative consiste à afficher un clavier virtuel sur l'écran de la TV. Plusieurs techniques et configurations de claviers sont envisageables. Par exemple avec *The Numeric-keypad Typer* [Ingmarsson *et al.*, 2004], l'alphabet présenté sur l'écran est divisé en 9 zones disposées matriciellement et identiquement aux touches numériques sur la télécommande (Figure 3.2). De la même manière, chaque zone contient neuf lettres ou symboles. Pour sélectionner un caractère, il suffit d'appuyer deux fois sur une touche numérique : d'abord pour sélectionner la zone puis la lettre dans la zone.

1	2	3
4	5	6
7	8	9

a	b	c	j	k	l	s	t	u
d	e	f	m	n	o	v	w	x
g	h	i	p	q	r	y	z	à
ä	ö	½	/	(	)	1	2	3
!	"	#	.	:	;	4	5	6
@	%	&	<	>	,	7	8	9
=	-	_		\	~	:	0	
~	£	\$	^	*	'			
[	{		u	:				SH

FIGURE 3.2: TNT [Ingmarsson *et al.*, 2004]. Correspondance entre le clavier numérique et les caractères.

Marshall *et al.* [2001] comparent six autres techniques et configurations de saisie de texte avec la télécommande. Notons que les auteurs affirment que la meilleure solution pour saisir des données alphanumériques reste le clavier physique classique. C'est sans doute pourquoi certaines télécommandes incluent un clavier complet auquel on accède en le dépliant ou en retournant la télécommande comme dans les travaux de Berglund et Johansson [2004] ou pour Boxee<sup>1</sup> (Figure 3.3).

### 3.2 TYPES DE TÉLÉCOMMANDES

Bien que la télécommande puisse être totalement dédiée au contrôle de la TV, la démocratisation des smartphones et surtout des tablettes tactiles alimente la tendance à consommer les produits télévisuels différemment. Par

1. <http://www.boxee.tv>



FIGURE 3.3: Télécommande Boxee équipée d'un clavier sur la face arrière.



FIGURE 3.4: Un exemple de télécommande doté d'un touchpad : la télécommande Panasonic Viera.

ailleurs, d'autres objets qui accompagnent l'activité télévisuelle, comme les magazines de TV, peuvent aider à accéder aux services plus facilement. Dans cette catégorisation des télécommandes, nous présentons deux groupes : les télécommandes dédiées et les télécommandes non dédiées au contrôle de la TV.

### 3.2.1 Télécommandes dédiées

Les télécommandes dédiées prennent plusieurs formes que nous séparons en trois groupes : les télécommandes traditionnelles augmentées, les télécommandes tangibles et les télécommandes dématérialisées.

#### 3.2.1.1 Télécommandes traditionnelles augmentées

L'exploitation de nouveaux capteurs et actuateurs permet d'augmenter les capacités interactionnelles de la télécommande.

Par exemple, la télécommande peut être équipée d'un touchpad [Enns et MacKenzie, 1998; Choi *et al.*, 2011], d'un joystick [Ishiyama et Yano, 2000; O'Modhain et Oakley, 2004] ou d'un trackball [MacKenzie et Jusoh, 2001]

pour améliorer le pointage et la sélection d'éléments graphiques sur les interfaces affichées sur l'écran, lesquelles sont inspirées des interfaces d'ordinateur de bureau [Ishiyama et Yano, 2000]. Elles permettent notamment de contrôler un pointeur à l'écran ou d'exécuter des commandes en traçant des symboles sur une surface [Enns et MacKenzie, 1998; Choi *et al.*, 2011] ou en l'air [Aoki *et al.*, 2012].

Les touchpads sont plus ou moins sophistiqués (Figure 3.4). Alors que Enns et MacKenzie [1998] utilisent un touchpad qui détecte seulement la position et le contact du doigt avec la surface, Choi *et al.* [2011] étendent les possibilités en fournissant la localisation en trois dimensions du pouce au dessus et un capteur de pression sous le touchpad.

Des composants capables de détecter l'orientation et le mouvement sont maintenant facilement intégrables aux télécommandes ceci afin d'exploiter davantage de degrés de liberté à l'interaction comme l'illustrent les télécommandes de Movea<sup>2</sup> ou la télécommande Wiimote de Nintendo<sup>3</sup> (Figure 3.5).

Les accéléromètres sont capables de mesurer l'accélération linéaire d'un dispositif de une jusqu'à trois dimensions. Ces mesures peuvent alors être interprétées pour déterminer la vitesse ou le déplacement linéaire et l'inclinaison d'un dispositif mobile. Il est alors déjà possible de quantifier les mouvements de translations et certains mouvement de rotation. Cependant, lorsque les axes du dispositif sont perpendiculaires au vecteur de la gravité, la mesure rapportée par les acéléromètres sur ces axes est nulle. C'est sans doute pourquoi dans *GestText* où une Wiimote permet de saisir du texte par le geste, Jones *et al.* [2010] ne considèrent que des gestes de tangage et de roulis (Figure 5.1).



FIGURE 3.5: Un exemple de télécommande augmentée : les possibilités de la Wiimote.

2. <http://www.movea.com/applications/interactive-tv>

3. <http://www.nintendo.com/wii/what-is-wii#/controls>

Des capteurs gyroscopiques peuvent compléter les capteurs accélérométriques. Ils mesurent la vitesse angulaire et fournissent des informations sur les mouvements de rotation du dispositif dans lequel ils sont logés. De nombreuses télécommandes en sont déjà équipées telles que les télécommandes de la Freebox intégrant les technologies de Movea ou encore le Wiimotion plus, aujourd'hui intégré dans les Wiimote, qui offre encore plus de précision pour le contrôle par mouvements [Wingrave *et al.*, 2010]. Contrairement aux accéléromètres, les gyroscopes ne sont pas affectés par la gravité mais seulement par le mouvement de rotation. De ce fait, ils peuvent être utilisés en combinaison avec les accéléromètres pour décomposer le mouvement par ses composantes translationnelles et rotationnelles.

Cependant, l'intégration par rapport au temps des mesures données par ces capteurs pour obtenir la vitesse, la position et l'orientation du dispositif conduisent à une dérive qui peut devenir importante. C'est pourquoi l'agrégation de ces données requiert des algorithmes de filtrage plus ou moins complexes avec une réponse rapide comme les filtres de Kalman [Wingrave *et al.*, 2010] ou plus récemment le filtre 1f [Casiez *et al.*, 2012].

La télécommande peut également être accompagnée d'un microphone offrant la possibilité au téléspectateur d'exprimer ses requêtes oralement [Wittenburg *et al.*, 2006]. Quelques modèles commerciaux exploitent déjà cette modalité comme les télécommandes de la société Amulet<sup>4</sup> ou Ruwido<sup>5</sup> (Figure 3.6). Comme nous le verrons plus tard, un avantage est la flexibilité descriptive des commandes vocales. En un seul énoncé, il est possible théoriquement d'exécuter des commandes complexes comme planifier l'enregistrement d'un programme à une date et un horaire précis [Berglund et Johansson, 2004].

La télécommande peut aussi transmettre de l'information au téléspectateur via des retours haptiques grâce à des moteurs de vibration [O'Modhrain et Oakley, 2003] ou encore des cellules brailles [Tahir *et al.*, 2008]. Le retour haptique étend l'expérience télévisuelle en accentuant l'effet d'immersion dans le programme [O'Modhrain et Oakley, 2003] ou véhicule des informations additionnelles sur les chaînes comme leur catégorie ou les contenus favoris [Tahir *et al.*, 2008].

---

4. <http://www.amuletdevices.com/index.php/products.html>

5. <http://www.ruwido.com/products/r117>



FIGURE 3.6: Un exemple de télécommande vocale : la télécommande Ruwido R117.

### 3.2.1.2 Télécommandes tangibles

La télécommande peut être conçue avec des formes particulières en s'appuyant sur des métaphores d'interaction et un couplage fort entre la manipulation de la télécommande et la présentation des données à l'écran.

Concevoir des télécommandes ayant une forme élémentaire permet de tirer partie de la faculté de la main à saisir et manipuler les objets [Ferscha *et al.*, 2007] (Figure 3.7). Par ailleurs, la correspondance entre la forme de la télécommande et l'interface graphique sur la TV facilite la prise en main des techniques d'interaction.

L'utilisation de formes simples a été explorée dans [Block *et al.*, 2004; Ferscha *et al.*, 2007; Kimman *et al.*, 2011; Tseng et Kolko, 2005]. Un cube peut par exemple être posé sur n'importe quelle surface horizontale et être facilement manipulé : les opérations pour ramasser, poser, pivoter, translater un cube sont intuitives [Block *et al.*, 2004; Ferscha *et al.*, 2007]. Partant de ce principe, Block *et al.* [2004] ont développé une technique permettant de sélectionner et visualiser un ensemble de chaînes à l'écran en manipulant le cube. Le vocabulaire gestuel décrit par les auteurs inclut les rotations du cube, secouer le cube, ou poser ce dernier dans une certaine position pour mettre le système télévisuel en veille. La combinaison d'un appui sur un bouton situé sur le cube et des différentes rotations permet d'augmenter sensiblement les possibilités. Ferscha *et al.* [2007] permettent aussi, de la même manière, de contrôler le volume et de naviguer à travers un menu. De forme cylindrique la télécommande dans [Kimman *et al.*, 2011] permet d'atteindre facilement avec le pouce les quatre anneaux disposés verticalement sur sa surface (Figure 3.8). L'apparence sphérique de TRIBA [Tseng et Kolko, 2005] suggère qu'on la manipule au moyen de rotations. La forme plane et large de Rebo invite l'utilisateur à interagir avec des gestes linéaires sur sa surface [Kobayashi *et al.*, 2009]. Ce-





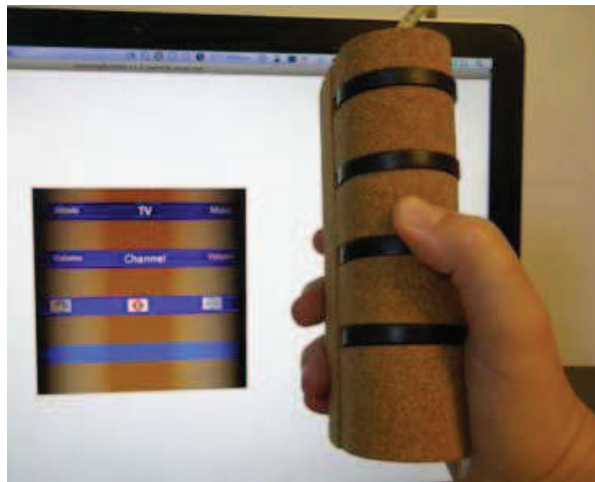
FIGURE 3.7: Un exemple de télécommande en forme de cube : le lien entre la forme de la télécommande et la représentation visuelle est très étroit [Block *et al.*, 2004].

pendant, la taille de *Rebo* nécessite qu'elle soit posée sur une surface stable et ne permet pas la manipulation avec une seule main.

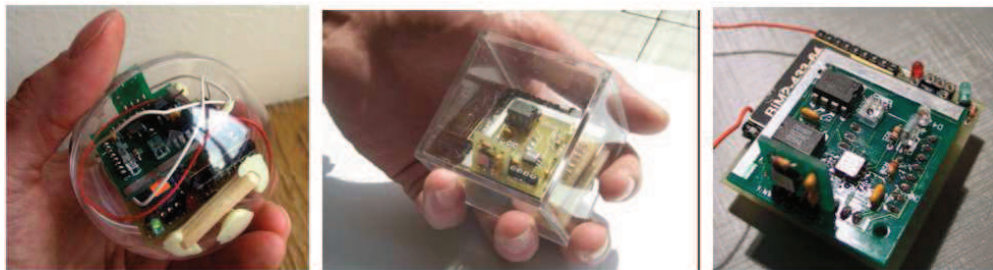
Des télécommandes de forme plus complexe comme *Peppermill* [Villar *et Hodges*, 2010] entraînent diverses affordances (Figure 3.10). La partie plane de *Peppermill* et la présence de boutons invitent l'utilisateur à tenir la télécommande alors que sa partie cylindrique suggère de visser ou dévisser pour sélectionner un élément dans une liste tout en rechargeant la télécommande en énergie. Enfin, la forme de *Tactimote*, proche d'un pistolet, permet de contrôler un joystick avec le pouce (à la place du chien) pour naviguer dans une représentation matricielle des chaînes TV [Tahir *et al.*, 2008]. La gâchette permet de valider avec le majeur tandis que l'index repose sur une cellule braille permettant de véhiculer des informations tactiles.

L'association entre la forme de la télécommande et la représentation logicielle à l'écran peut créer une connexion évidente entre le dispositif physique et l'interface numérique. *TRIBA* permet de naviguer à travers une collection de contenus multimédias représentée sous forme d'une composition sphé-



FIGURE 3.8: Spinning Remote [Kimman *et al.*, 2011].

rique [Tseng et Kolko, 2005]. Chaque chaîne est représentée par une épingle pouvant se dilater pour afficher un aperçu du contenu. Cette solution permet la visibilité de toutes les options offertes, même si l'intérêt est seulement porté sur un groupe particulier de chaînes.

FIGURE 3.9: Exploration de formes pour la manipulation de dispositifs (à gauche une sphère et au centre un cube ; à droite Smart-Its une plateforme de prototypage) [Block *et al.*, 2004].

De même, le prototype de Block *et al.* [2004] dont nous avons déjà parlé est associée à une représentation cubique de telle sorte que lorsque que le cube physique est soumis à une rotation, le cube virtuel restitue cette rotation sur la TV et permet de visionner l'aperçu de trois chaînes différentes simultanément sur les faces visibles (Figure 3.7). Si le cube est posé et immobile, la chaîne sur la face la plus visible du cube virtuel s'élargit pour couvrir tout l'écran. Visser ou dévisser *Perpermill* permet d'actionner un carrousel virtuel présentant les différents contenus multimédia [Villar et Hodges, 2010]. Enfin, dans [Kimman *et al.*, 2011] la représentation visuelle se présente sous la forme

de quatre menus, chacun étant associé à chaque anneau de la télécommande permettant d'interagir sans avoir à détourner le regard de l'écran de TV (Figure 3.8).



FIGURE 3.10: Peppermill [Villar et Hodges, 2010].

D'une autre manière, les gestes peuvent aussi être liés à la représentation graphique. Les gestes réalisés sur *Rebo* permettent d'engager une commande et de la valider [Kobayashi *et al.*, 2009]. Par exemple, pour changer de chaîne, l'utilisateur effectue un trait horizontal sur la télécommande (Figure 3.11). Au fur et à mesure qu'il réalise son geste, un aperçu de la chaîne suivante apparaît et pousse la chaîne actuelle vers l'extérieur de l'écran. Si le geste s'arrête, la commande avorte et la chaîne n'est pas modifiée. Ce système d'exécution partielle, appelé *action slopping* par ses auteurs, permet de prendre conscience des fonctionnalités existantes et introduit une certaine tolérance à l'erreur, indisponible avec une télécommande traditionnelle. D'autres exemples sont présentés dans [Kobayashi *et al.*, 2009].

### 3.2.1.3 Télécommandes dématérialisées

La conception de techniques d'interaction alternatives est nécessaire pour palier aux inconvénients de la télécommande physique. Par exemple, la télécommande n'est pas toujours à portée de main [Freeman T. et Craig D., 1995], l'utilisateur pouvant être plus ou moins éloigné du canapé ou engagé dans une autre tâche. Elle peut de plus avoir été égarée [Freeman T. et Craig D., 1995; Tsekleves *et al.*, 2007]. Enfin certains utilisateurs peuvent avoir du mal à utiliser les télécommandes traditionnelles, notamment s'ils sont mal voyants [Berglund et Johansson, 2004].



FIGURE 3.11: *Rebo* : un geste sur le corps de la télécommande permet de changer de chaîne. [Kobayashi *et al.*, 2009].

Plusieurs solutions sont envisageables pour contrôler la TV sans avoir à manipuler de dispositif physique. Des solutions basées sur des systèmes de capture vidéo couplés à un moteur de reconnaissance de gestes sont explorées dans [Freeman T. et Craig D., 1995; Bailly *et al.*, 2011; Yoo *et al.*, 2011; Vatavu, 2012; Ren et O'Neill, 2012]. La popularisation du Kinect de Microsoft ouvre de nombreuses perspectives suivant cette approche [Bailly *et al.*, 2011; Yoo *et al.*, 2011; Vatavu, 2012; Ren et O'Neill, 2012]. Le Kinect est un périphérique initialement prévu pour la console de jeux vidéos de Microsoft (Figure 3.12). Ce périphérique est doté d'une caméra couleur, d'une caméra de profondeur et de microphones. La disponibilité de pilotes et d'une interface de programmation libre (OpenNI<sup>6</sup>) ont contribué aux prémices de son succès. Aujourd'hui, il est même possible de développer des applications avec le kit de développement officiel dédié<sup>7</sup>.



FIGURE 3.12: Kinect for Windows.

---

6. <http://www.openni.org/>

7. <http://www.microsoft.com/en-us/kinectforwindows>

Grâce à ce périphérique, l'idée de transposer l'interaction main-libre du laboratoire au salon inspirent les chercheurs. Par exemple, [Bailly et al. \[2011\]](#) comparent trois techniques de menu basés sur les gestes pour sélectionner des commandes pour la TV. Suivant une autre stratégie, [Vatavu \[2012\]](#) rapporte une étude de conception de vocabulaire gestuel pour la TV par les utilisateurs. Des travaux s'intéressent également à la sélection et la manipulation d'éléments dans un espace en trois dimensions [[Yoo et al., 2011](#)] ou la sélection d'éléments dans un menu pour les TVs dotés d'écran affichant de la 3D stéréoscopique, une nouvelle génération de TV [[Ren et O'Neill, 2012](#)].

Comme pour les télécommandes augmentées, l'utilisation de la voix est une autre alternative. Des systèmes proposent de discuter directement avec un agent conversationnel. Ces agents peuvent prendre la forme d'avatars virtuels affichés sur l'écran de la TV [[Ugarte et al., 2007](#)] ou d'agents physiques telle une peluche [[Goto et al., 2003](#)].

Enfin la joaillerie numérique permet également de s'abstraire de télécommande physique [[Perrault et al., 2011](#)]. Embarquant des capteurs d'orientations, *Actual Remote Control* est un bracelet qui transmet des informations sur le mouvement de l'avant-bras dont le geste permet de sélectionner des commandes sur l'interface affichée par l'écran de la TV [[Lee et al., 2009](#)]. De la même manière, [Zhang et al. \[2011\]](#) traquent le mouvement de l'index de la main grâce aux capteurs inclus dans une bague. Également équipée d'un microphone la bague peut aussi détecter le contact du doigt avec n'importe quelle surface grâce à la conduction du son par les os.

### 3.2.2 Télécommandes non dédiées

Comme évoqué auparavant, la tendance à utiliser les tablettes tactiles devant la TV ne cesse de se renforcer. Des travaux explorent la possibilité d'enrichir les usages liés à la TV. D'autres travaux s'appuient sur les connaissances des téléspectateurs à utiliser des objets en relation avec la TV pour simplifier l'utilisation des services. Nous décrivons ici ces deux approches.

#### 3.2.2.1 Télécommandes d'accompagnement (*Companion devices*)

Les dispositifs mobiles comme les "smartphones" et les tablettes tactiles, dont le nombre ne cesse de croître, ont le double avantage de permettre un accès facile aux infrastructures réseau et de disposer d'une surface d'affichage suffisamment grande pour véhiculer une quantité d'information importante. Ils sont de plus généralement équipés d'un écran tactile et de divers capteurs

comme des accéléromètres, des gyroscopes et un microphone.

L'utilisation de dispositifs mobiles d'appoint pendant l'activité télévisuelle est une tendance importante. L'agence d'étude et d'information Nielsen rapporte un sondage sur l'utilisation d'un smartphone ou d'une tablette tactile en regardant la TV<sup>8</sup> pour le dernier quadrimestre 2011 (Figure 3.13). L'agence annonce qu'aux Etats-Unis, 88% de possesseurs de tablette et 86% de possesseurs de smartphones utilisaient leurs terminaux au moins une fois par mois en regardant la TV. Au Royaume-Uni, cette proportion est de 80% pour les possesseurs de tablette et de 78% pour les possesseurs de smartphone.

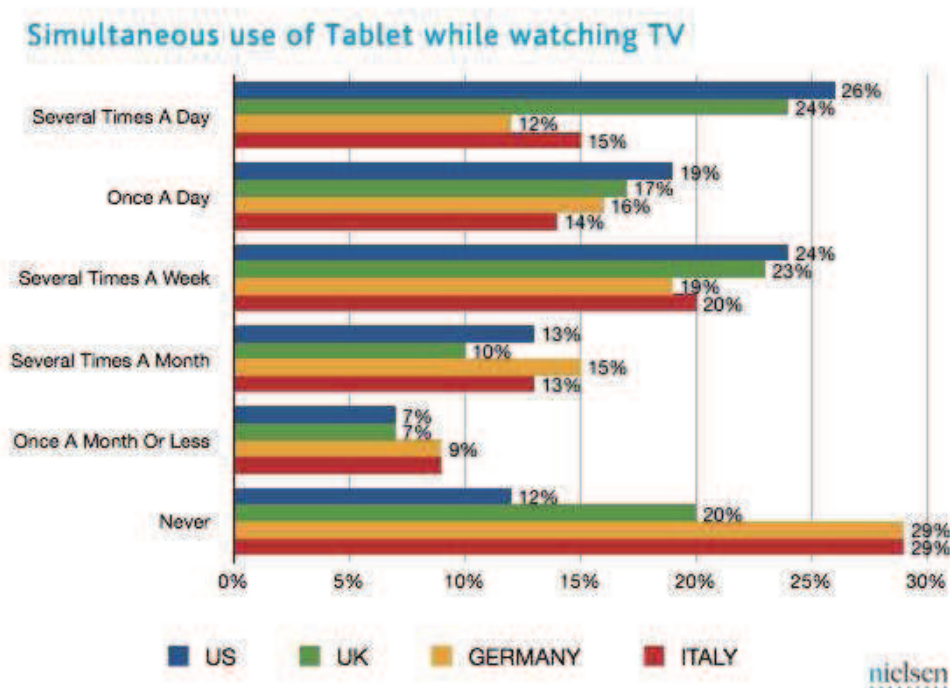


FIGURE 3.13: Utilisation de la tablette pendant l'activité télévisuelle (source : Institut de sondage Nielsen, Mai 2012).

De par la taille de leur écran ces dispositifs peuvent aussi servir d'écran auxiliaire [Cesar *et al.*, 2008; Cruickshank *et al.*, 2007; Robertson *et al.*, 1996] ceci permettant d'éviter l'apparition d'informations sur l'écran de la TV propres à gêner les autres utilisateurs. Cette solution permet de contrôler, de partager, de commenter ou d'enrichir le contenu diffusé, particulièrement dans le cas des tablettes dont la taille facilite l'entrée de texte. Cette solution permet aussi

8. <http://bit.ly/ZvsjWf>

de personnaliser l'interaction avec le système en adaptant l'interface en fonction de l'utilisateur, par exemple, pour contrôler l'accès aux contenus [Tsekleves *et al.*, 2007] ou pour personnaliser les contenus favoris [Cruickshank *et al.*, 2007]. Cette approche intéresse également les fournisseurs de contenus et de décodeurs numériques qui travaillent sur des solutions applicatives à l'instar de Technicolor avec le projet Quaero PVAA++<sup>9</sup>. Une architecture pour réaliser un tel système est détaillée dans [Cesar *et al.*, 2008].

Le dispositif mobile peut aussi jouer le même rôle qu'une télécommande augmentée dans la mesure où il est généralement équipé de capteurs en plus de son écran tactile. Par exemple le système multimodal de Turunen *et al.* [2009] permet de naviguer ou zoomer en effectuant des gestes dans l'air avec le téléphone ou d'activer des commandes par la voix ou par contact du téléphone sur des objets comportant des étiquettes RFID. Le retour haptique est utilisé pour avvertir l'utilisateur du bon déroulement, ou non, des opérations sous forme d'icônes haptiques associées à des comportements vibratoires différents.

### 3.2.2.2 Objets du quotidien augmentés



FIGURE 3.14: Un pouf augmenté pour contrôler la télévision. [Schmidt *et al.*, 2004]

Pour faciliter l'apprentissage de nouvelles techniques ou technologies interactives, la réutilisation des objets du quotidien tire avantage de la familia-

9. <http://bit.ly/QuaeroPVAAPlusPlus>



rité avec ces objets. Basé sur une étude des routines quotidiennes au domicile, *Paper Remote* combine un stylo numérique et un guide papier des programmes télévisés pour spécifier les programmes à visionner ou à enregistrer (Figure 3.15). Ce système introduit donc une nouvelle technologie par le biais d'un artefact lié à la TV. L'interaction avec le papier a été aussi investiguée dans [Hess *et al.*, 2008] pour personnaliser la télécommande pour contrôler un environnement multimédia et dans le cadre de l'enseignement électronique [Lima et Nevado, 2007]. Via un lecteur et des codes barres imprimés sur le cours, l'étudiant peut jouer des vidéos liées aux documents papier sur la TV.



FIGURE 3.15: Paper Remote : un guide de programme TV augmenté d'icônes cliquables avec un stylo numérique. [Berglund *et al.*, 2006]

Certains travaux s'intéressent davantage aux relations avec les objets et à l'expérience utilisateur [Lee *et al.*, 2007; Nack *et al.*, 2007; Schmidt *et al.*, 2004; Vatavu et Pentiuc, 2008]. Dans [Lee *et al.*, 2007], lancer un ballon en mousse permet de modifier le contenu diffusé par celle-ci en fonction de l'état émotionnel (Figure 3.16). Dans [Nack *et al.*, 2007], les auteurs envisagent d'utiliser un coussin pour contrôler la TV, en particulier lorsque la télécommande est hors d'atteinte. De plus, la relation intime entre le téléspectateur et le coussin permet de surveiller ses données biométriques. L'analyse de ces données peut permettre de suggérer des activités à l'utilisateur tout en personnalisant l'interface sur la surface du coussin pour cette activité, comme par exemple un jeu. Dans le même esprit ludique, Schmidt *et al.* [2004] proposent d'utiliser un pouf comme télécommande (Figure 3.14). Enfin, la table basse du salon peut aussi être utilisée comme surface de contrôle [Vatavu et Pentiuc, 2008]. Celle-ci a de plus le mérite d'être partageable par tous les utilisateurs situés autour de la table, lesquels interagissent au moyen de gestes.



FIGURE 3.16: Un ballon capable de sentir les émotions. [Lee *et al.*, 2007]

### 3.3 DIMENSION INTERACTIONNELLE

Nous avons réalisé une première classification des télécommandes qui s'appuie sur leur type et leurs usages vis à vis du contexte de la TV. Néanmoins, ces télécommandes partagent des caractéristiques comme la modalité qu'elles utilisent en entrée ou encore le type d'information qu'elles transmettent au téléspectateur. Nous présentons ces caractéristiques.

#### 3.3.1 Entrée

##### 3.3.1.1 Les boutons

En terme de modalités, l'interaction sur une télécommande traditionnelle se limite à un appui discret sur une touche à sa surface. Le geste peut être répétitif et fastidieux lorsque le système offre de nombreuses possibilités, comme précédemment expliqué. Contrairement aux ordinateurs personnels, ces interfaces ne fournissent pas d'équivalent de la souris avec un contrôle continu suivant deux dimensions. Ceci interdit tout autant le pointage vers des objets interactifs affichés à l'écran que l'interaction gestuelle via des techniques comme les Marking menus [Kurtenbach et Buxton, 1991]. Notons, cependant, que la localisation de la touche sur la télécommande peut être porteuse de sens comme dans *The Numeric-keypad Typer* où elle définit un groupe de caractère particulier [Ingmarsson *et al.*, 2004]. Le temps d'appui sur la touche est également utilisé pour ajouter une fonctionnalité au bouton comme la sélection d'une chaîne dont le numéro est composé de plusieurs chiffres. L'appui prolongé sur une touche numérique permet de valider un



chiffre et offre la possibilité de saisir le chiffre suivant.

### 3.3.1.2 Les gestes sur une surface

Les gestes sur surface peuvent prendre plusieurs formes dans le contexte de la TV. Enns et MacKenzie [1998] proposent d'utiliser des gestes d'un seul trait représentant des chiffres ou des caractères alphabétiques effectués avec le pouce sur une télécommande pour accéder aux commandes basiques de la TV. En plus de 10 chiffres, leur système reconnaît des commandes pour éteindre, jouer une vidéo, avancer ou rembobiner rapidement en dessinant respectivement les caractères x, w, p, f ou r. En outre, Choi *et al.* [2011] évaluent leur prototype avec une tâche de flicking au pouce pour choisir des chaînes disposées en liste horizontale contrôlables par un slider affiché sur l'écran de TV. Ils évaluent également une tâche d'entrée de texte avec un clavier au format *Multitap*.

Choi *et al.* [2011] montrent l'avantage du retour visuel sur la TV lorsque le pouce se déplace au-dessus de la surface interactive notamment pour faciliter la réalisation des gestes numériques sur le trackpad de la télécommande. Ils proposent une étude de comparaison de l'usage d'un trackpad au pouce sur une télécommande avec et sans retour visuel sur la TV lorsque le pouce est au-dessus de la surface. Bien que les résultats pour les tâches de flicking ou d'entrée de texte ne soient pas meilleurs, ils rapportent que les utilisateurs perçoivent moins de fatigue et un retour d'information plus rapide lorsque le retour visuel est présent. Ce résultat peut paraître anodin de prime abord, mais comme discuté auparavant, dans un contexte où l'efficacité n'est pas la seule priorité, il reste intéressant et mérite une attention particulière.

Les dispositifs mobiles disposent d'un écran tactile. Ils accueillent de nombreuses applications<sup>10</sup> permettant de remplacer les télécommandes traditionnelles et les boutons physiques par des widgets sur lesquels le téléspectateur appuie pour lancer une commande. Cette alternative présente un inconvénient majeur. Les propriétés physiques des boutons de la télécommande comme leur forme ou leur texture ne sont pas préservées par l'écran tactile et, de fait, ils accaparent entièrement l'attention visuelle pour exécuter les commandes.

La taille de l'affichage permet selon les dispositifs d'afficher un nombre important d'informations ou de reproduire l'interface disponible sur la TV (Figure 3.17). Ces dispositifs facilitent la sélection directe d'une commande

---

10. <http://gizmodo.com/5982909/5-apps-to-turn-your-phone-into-a-universal-remote>

par opposition à la navigation séquentielle imposée par la télécommande traditionnelle. Naturellement, ces dispositifs peuvent également supporter les gestes disponibles sur trackpad discutés précédemment.



FIGURE 3.17: Une application de guide électronique des programmes et d'accès aux chaînes sur un dispositif mobile [Cruickshank *et al.*, 2007]

### 3.3.1.3 Les gestes dans un espace à 3 dimensions

Les gestes en 3 dimensions peuvent être exploités pour la navigation [Gurrin *et al.*, 2010; Lee *et al.*, 2009; Zigelbaum *et al.*, 2010], l'exécution de commandes [Zigelbaum *et al.*, 2010; Bailly *et al.*, 2011], la saisie de texte [Aoki *et al.*, 2010; Jones *et al.*, 2010] et le pointage [Freeman T. et Craig D., 1995; Ishiyama et Yano, 2000; MacKenzie et Jusoh, 2001].

Gurrin *et al.* [2010] utilisent des gestes dans l'espace pour naviguer dans une collection de souvenirs numériques. De simples gestes avec la télécommande permettent de se déplacer horizontalement d'événement en événement et verticalement de jour en jour. Cependant, les auteurs ne décrivent pas précisément quels types de gestes sont associés à ces actions. Par contre, lorsqu'un événement est sélectionné, les téléspectateurs peuvent contrôler la vitesse de diffusion des souvenirs en inclinant la télécommande de gauche à droite à la manière d'un cadran circulaire téléphonique ou d'une poignée de porte. Les auteurs insistent sur l'utilisation de gestes simples et peu nombreux afin de minimiser l'effort cognitif et le temps d'apprentissage pour en fin de compte maximiser le plaisir des téléspectateurs.



FIGURE 3.18: G-Stalt : L'utilisateur manipule des vidéos à l'aide de gestes main-libre. [Zigelbaum *et al.*, 2010]

*G-Stalt* propose des gestes permettant d'activer des modes, de contrôler avec précision des paramètres continus ou d'exécuter des commandes. C'est un système qui permet de naviguer à travers une collection de média vidéos dans un espace graphique à trois dimensions [Zigelbaum *et al.*, 2010] (Figure 3.18). Il est possible de jouer les vidéos, rechercher des informations, les réorganiser en fonction de leurs méta-données ou encore les déplacer dans l'espace. Par exemple, pincer avec les deux mains permet de déclencher le mode de contrôle de l'espace. Par la suite, l'utilisateur déplace ses deux mains pour réaliser des translations ou des rotations de l'espace. Un autre exemple est l'activation du mode d'agencement des vidéos lorsque l'utilisateur se touche la tête avec une main. L'autre main est utilisée pour dessiner un circuit en 3 dimensions qui servira de guide pour l'alignement spatial de la collection de vidéos. Un exemple d'exécution de commande est la lecture d'une vidéo qui se déclenche en montrant sa main au système la paume vers le haut.

Comme vu précédemment, *Actual Remote* offre la navigation à l'intérieur d'un menu grâce aux gestes de l'avant-bras [Lee *et al.*, 2009]. Chaque geste de  $10^\circ$  dans une direction horizontale ou verticale est converti en un événement discret dans la même direction. Par exemple, un mouvement de l'avant bras de  $10^\circ$  vers la droite produit une commande qui permet de déplacer le sélecteur vers la droite d'une unité de déplacement. Un mouvement de  $70^\circ$

est équivalent à 7 commandes vers la droite.



FIGURE 3.19: Finger Count : Montrer ses mains et un nombre de doigt au système permet de sélectionner un élément dans un menu affiché à distance sur la TV [Bailly *et al.*, 2011]

Bailly *et al.* [2011] proposent d'utiliser des gestes basés sur des directions ou sur le nombre de doigts levés de la main pour contrôler écran à distance (Figure 3.19). A la manière des *Marking menus* sur surface interactive [Kurtenbach *et al.*, 1993], les directions des gestes permettent d'exécuter une commande dans un menu ayant une représentation circulaire. Pour augmenter le nombre de commandes accessibles, les auteurs s'inspirent des *Multi-strokes Marking Menus* proposés par Zhao et Balakrishnan [2004]. La seconde technique est une adaptation dans un espace en trois dimensions du menu *Finger Count* où le nombre de doigts en contact sur la surface interactive permet de sélectionner un élément d'un menu [Bailly *et al.*, 2010]. Dans cette étude, le système comptabilise le nombre de doigts levés pour chaque main lorsque l'utilisateur présente ses mains paumes face à l'écran. La comparaison de ces techniques de menus montrent que la technique inspirée du menu *FingerCount* est cognitivement plus demandeuse que celle inspirée des *MarkingMenus*, un détail qui peut avoir son importance dans l'environnement domestique [Drucker *et al.*, 2002; Gurrin *et al.*, 2010].


*GesText* permet d'entrer du texte en effectuant uniquement des gestes d'inclinaison de la télécommande pour sélectionner un caractère dans différents arrangements de claviers virtuels [Jones *et al.*, 2010]. Les auteurs étudient notamment la capacité des utilisateurs à réaliser une inclinaison verticale (pour une sélection verticale) ou une inclinaison par rapport à l'axe longitudinal de la télécommande (pour une sélection horizontale). La validation de la sélection d'un caractère est effectuée par une fixation du geste pendant une période de temps donnée. Les auteurs s'intéressent également au rapport entre l'amplitude du geste et le niveau de sélection, et à la sélection en direction diagonale. Les résultats montrent que les utilisateurs arrivent difficilement à effectuer des sélections correctes après deux niveaux de profondeur et à combiner simultanément les deux types d'inclinaison pour sélectionner des éléments diagonaux.

Dans *Twist&Tap*, la saisie de kanas (caractères japonais) se réalise en combinant un geste de rotation sur l'axe de la profondeur de la télécommande et l'appui sur une touche du pavé directionnel [Aoki *et al.*, 2010]. Le degré d'inclinaison de la télécommande est associé à une paire de consonnes qu'il est nécessaire de distinguer avec un appui des touches gauche ou droite du pavé directionnel. Les touches haut/bas de la télécommande permettent la sélection des voyelles pour finalement former une syllabe (Figure 3.20). Les auteurs suggèrent que la parallélisation de l'appui sur un bouton et le mouvement de la télécommande pour sélectionner une consonne pourrait être plus rapide qu'une sélection séquentielle.

Des gestes translationnels de la télécommande peuvent aussi permettre la sélection de caractères. *CubeKey* est une technique de saisie de texte basée sur un clavier avec une disposition cubique de longueur 3 [Shoemaker *et al.*, 2010]. Pour accéder à chacun des caractères, il suffit d'effectuer un mouvement de translation de la télécommande dans cet espace cubique (Figure 5.4). La validation d'un caractère est opérée par l'appui d'un bouton. Cette technique a la propriété d'être indépendante à la distance à l'écran. Shoemaker *et al.* [2010] définissent une technique dépendante à la distance comme une technique qui change l'action physique quand la distance entre l'utilisateur et l'écran change. Plus concrètement dans *CubeKey*, le mouvement du curseur de sélection est relatif au mouvement du dispositif et les gestes à réaliser pour sélectionner un caractère ne varient pas en fonction de la distance à l'écran. Cette propriété est également intéressante pour le contexte de la TV.

Enfin, les gestes dans l'espace permettent de contrôler un curseur sur l'écran [Freeman T. et Craig D., 1995]. Le prototype proposé par Freeman et Craig



			Left key					Right key				
			W	R	Y	M	H	N	T	S	K	Null
	0		わ	ら	や	ま	は	な	た	さ	か	あ
	1		を	り	ゆ	み	ひ	に	ち	し	き	い
	2		ん	る	よ	む	ふ	ぬ	つ	す	く	う
	3		,	れ	や	め	へ	ね	て	せ	け	え
	4		.	ろ	ゆ	も	ほ	の	と	そ	こ	お
	5		ー		よ				っ			あ
	6		*									い
	7		*									う
	8											え
	9											お

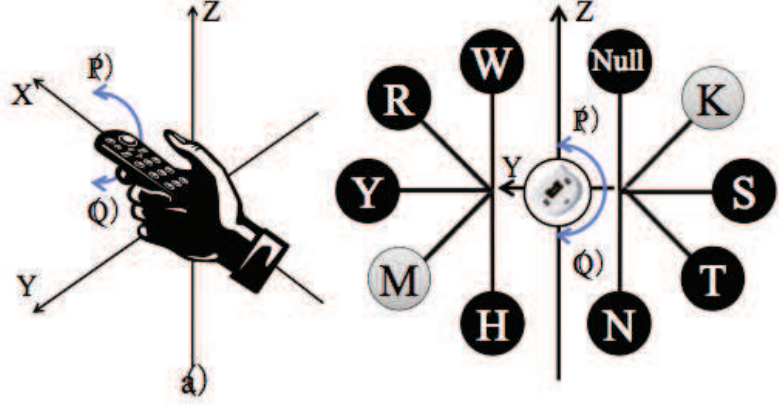
  


FIGURE 3.20: Twist&Tap : sélection de kanas en inclinant et appuyant sur les boutons de la télécommande [Aoki *et al.*, 2010].

propose de considérer la main comme une souris sans bouton. Lorsque le système est activé, une couche de widgets est ajoutée par-dessus le programme diffusé. Pour contrôler un de ces widgets, il suffit de déplacer sa main dessus et d'attendre 200 millisecondes pour activer la sélection ou exécuter une commande. Le déplacement de la main dans le plan frontal de l'utilisateur est directement mis en correspondance avec le curseur à l'écran. Les auteurs rapportent l'excitation des utilisateurs mais aussi, la fatigue physique liée à l'utilisation intensive du système.

Par ailleurs, le pointage dans l'air via un dispositif gyroscopique est comparé à d'autres techniques dans [Ishiyama et Yano, 2000] (touches directionnelles ou trackball) et [MacKenzie et Jusoh, 2001] (souris ou joystick). Ishiyama et Yano [2000] montrent que les temps de pointage ne dépendent pas de la distance entre l'utilisateur et l'écran mais du gain appliqué entre le dispositif d'entrée et le curseur. Par ailleurs, le dispositif de pointage gyroscopique obtient de meilleures performances que le dispositif doté d'un trackball et que

la télécommande munie de touches directionnelles.

En comparant deux techniques de pointage en l'air et deux techniques de pointage de bureau, MacKenzie et Jusoh montrent que les techniques de pointage en l'air sont moins performantes et moins facile à utiliser que leurs pendantes pour l'interaction au bureau [MacKenzie et Jusoh, 2001]. Un dispositif doté d'un joystick isométrique, un autre doté d'un gyroscope et une souris sont utilisés dans cette étude. Deux conditions sont évaluées : une condition où l'utilisateur est installé au bureau avec la souris ou le dispositif gyroscopique et une autre où il est debout à une distance de 1,5 mètres de l'écran avec le dispositif doté d'un joystick isométrique ou le dispositif gyroscopique. La comparaison entre l'utilisation des différents dispositifs montre que le temps moyen de sélection avec le dispositif gyroscopique est inférieur qu'avec le dispositif doté du joystick. Cependant, ce dispositif s'avère plus précis. Notons quand même que de nombreuses différences existent entre tous ces dispositifs et la manière de les opérer ce qui peut avoir une influence sur les résultats de cette étude.

#### 3.3.1.4 *Le langage naturel*

La reconnaissance du langage naturel constitue une autre possibilité pour exécuter les commandes basiques de la TV [Balchandran *et al.*, 2008; Ibrahim *et al.*, 2001] la navigation parmi les chaînes de TV [Balchandran *et al.*, 2008; Ibrahim *et al.*, 2001; Goto *et al.*, 2003], la recherche d'information [Tseng et Kolkko, 2005; Berglund et Johansson, 2004; Goto *et al.*, 2003]. Une des ambitions des commandes vocales est de faciliter l'utilisation des systèmes par des utilisateurs novices. Elles permettent aussi de créer des raccourcis descriptifs afin de réduire le nombre d'actions pour réaliser les tâches complexes [Berglund et Johansson, 2004; Ibrahim *et al.*, 2001].

Goto *et al.* [2003] se sont intéressés au type de langage que les téléspectateurs adopteraient avec un agent conversationnel pour la TV. En présentant un avatar aux participants, les auteurs proposent de converser avec l'agent qui peut comprendre tout ce qu'il entend (Figure 3.21). En réalité, un opérateur actionne les commandes et manipule l'interface aux fur et à mesures des requêtes des participants. Il rapportent que 83% des propos tenus par les participants concernent la TV et parmi ces propos, 89% incluent des mots appartenant à des catégories spécifiques comme : les titres des programmes, le genre de programme, les acteurs, les canaux de diffusion, l'heure de diffusion, et les commandes de la TV. Basé sur ces catégorisations, les auteurs proposent un système capable de sélectionner des chaînes, d'exécuter des



FIGURE 3.21: Un agent conversationnel pour la TV sous forme de peluche [Goto *et al.*, 2003].

commandes basiques de la TV, de présenter des informations sur le contenu des programmes, de planifier un enregistrement de programme à un horaire défini ou de naviguer de programme en programme via les informations qu'ils partagent. De plus, Balchandran *et al.* [2008] proposent des commandes vocales qui permettent le défilement ou la sélection d'un élément dans une liste.

Berglund et Johansson [2004] soulignent les avantages du langage naturel dans le contexte de la TV. D'abord, le langage naturel permet de spécifier une action et un objet et de décrire un objet avec plusieurs propriétés dans une unique commande. Il peut également se substituer aux techniques d'entrée de texte. Enfin, il permet d'accéder à des commandes qui ne sont pas affichées sur l'interface. Les auteurs proposent d'utiliser le langage naturel pour piloter un guide électronique des programmes. En particulier, le langage naturel est exploité pour fournir de la flexibilité pour la recherche de programme dans le guide. De la même manière *TRIBA* vu précédemment supporte le langage naturel pour rechercher des contenus en adéquation avec la requête du téléspectateur [Tseng et Kolko, 2005].

Cependant, le langage naturel ne possède pas que des avantages. Suite à l'évaluation de leur système chez les participants, Berglund et Johansson [2004] identifient les situations dans lesquelles la télécommande reste de mise. La première situation est lorsque l'action de parler peut déranger d'autres personnes où élever le volume sonore de l'environnement à outrance. La mo-



dalité vocale peut être aussi engagée dans une autre activité comme parler avec d'autres téléspectateurs. Parfois, il est difficile d'exprimer une requête et, dans ce cas, la télécommande est essentielle pour naviguer dans un menu. Par ailleurs, parler naturellement avec un agent conversationnel peut être difficile pour certains utilisateurs [Berglund et Johansson, 2004; Tseng et Kolko, 2005].

Le langage naturel propose de nombreux avantages pour faciliter l'accès aux commandes complexes ou exprimer des requêtes sophistiquées. Cependant, ses limites l'empêchent de remplacer totalement la télécommande [Berglund et Johansson, 2004; Ibrahim *et al.*, 2001].

### 3.3.2 Sortie

#### 3.3.2.1 Le retour haptique

Comme nous l'avons déjà dit à la section 3.2.1.1, le retour haptique sur la télécommande peut servir à véhiculer des informations additionnelles au contenu télévisuel [Tahir *et al.*, 2008] ou augmenter l'engagement du téléspectateur avec le contenu télévisé [O'Modhain et Oakley, 2003, 2004].



FIGURE 3.22: *Tactimote* : une cellule braille permet de transmettre de l'information par l'index. [Tahir *et al.*, 2008]

Équipée d'une cellule braille, *Tactimote* est une télécommande qui permet une navigation rapide dans une liste de chaînes de TV [Tahir *et al.*, 2008] (Figure 3.22). Les auteurs considèrent 7 patrons statiques et clignotants. L'ensemble des chaînes accessibles est divisé en 7 catégories, chacune correspondant à un patron statique. Lorsque l'utilisateur navigue d'une catégorie à une autre le patron transmis par la cellule braille est modifié. Si une chaîne favorite est sélectionnée, la cellule braille est actualisée avec le patron clignotant

correspondant à la chaîne. Les auteurs rapportent qu'il est possible, en mode expert, d'identifier la chaîne favorite uniquement par l'information transmise par la cellule braille.

Le retour tactile est aussi utilisé pour avertir l'utilisateur du bon déroulement des opérations. Dans une tâche de lecture d'étiquettes RFID avec un téléphone portable pour sélectionner des commandes, [Turunen et al. \[2009\]](#) mettent en place deux icônes haptiques matérialisées par deux comportements vibratoires différents. Alors que le premier patron assure que le contact entre le téléphone et l'étiquette s'est déroulé correctement, le second signale toute sorte d'erreur.

[O'Modhrain et Oakley \[2003, 2004\]](#) exploitent le retour haptique pour augmenter le niveau subjectif d'immersion et d'implication des téléspectateurs et l'interactivité avec les programmes. Ils explorent deux types de retour, le retour haptique pendant une exploration active du toucher et le retour haptique avec une réception passive de la main inerte. Dans la situation d'exploration active, le retour est généré en réponse aux actions du téléspectateur. Un exemple d'application d'exploration active est la transmission du retour haptique pendant que le téléspectateur contrôle et ressent le mouvement d'une abeille qui vole. Le retour passif se produit lorsque par exemple le téléspectateur ressent les forces exercées sur un objet en chute libre. Notons au passage que ce procédé est largement exploité dans les jeux vidéos.

### 3.3.2.2 *Le retour auditif*

A notre connaissance, il n'y a pas de travaux s'intéressant effectivement au retour auditif de la télécommande, exceptés les interfaces vocales et les agents conversationnels qui peuvent répondre vocalement à chaque requête du téléspectateur [[Goto et al., 2003](#)]. Ceci peut s'expliquer par le fait que la TV est déjà équipée de haut-parleurs qui permettent de véhiculer des informations sonores. De plus, le niveau sonore déjà important de la TV lorsqu'elle est allumée permet difficilement la transmission de messages audio via la télécommande. Par ailleurs, nous avons vu précédemment que l'activité télévisuelle peut être partagée par plusieurs personnes et par d'autres activités. Ces situations potentiellement bruyantes peuvent limiter l'efficacité d'un tel canal d'information.

### 3.3.2.3 *Le retour visuel*

Le retour visuel passe principalement par l'écran de la TV où la plupart des interfaces et informations sont affichées. Cependant, certains dispositifs équipés d'un écran permettent de véhiculer des informations visuelles. C'est en particulier le cas des tablettes interactives ou des smartphones qui peuvent aussi de par la taille et la résolution de leur écran servir d'écran auxiliaire [Cesar *et al.*, 2008; Cruickshank *et al.*, 2007]. Cela permet d'éviter l'apparition d'informations sur l'écran de la TV susceptibles gêner les autres utilisateurs comme déjà abordé dans la section 3.2.2.1 de ce chapitre.

Cruickshank *et al.* [2007] proposent d'utiliser un second écran pour résoudre le conflit entre l'affichage des informations utiles sur l'écran et l'occultation du programme en cours de diffusion, exprimée par les familles possédant une TV interactive. Cette approche permet de libérer l'écran principal de l'affichage des éléments de l'interface graphique. Elle offre aussi l'opportunité de personnaliser l'expérience de la TV en reconfigurant dynamiquement l'interface utilisateur offerte au téléspectateur comme afficher les favoris pour un utilisateur en particulier. Un avantage supplémentaire mentionné par les auteurs est la possibilité de consulter des informations en respectant les autres téléspectateurs dont l'attention visuelle est dirigée vers l'écran principal. Par exemple, il est alors possible de consulter le guide de programme électronique [Cruickshank *et al.*, 2007] ou tout autre service en relation avec la TV [Cesar *et al.*, 2008].

L'inverse peut aussi se produire comme dans [Robertson *et al.*, 1996] où les auteurs proposent une application de location et d'achat de biens immobiliers dans laquelle l'écran de TV est utilisé pour montrer les photos des appartements alors que les informations principales sont affichées sur l'écran d'un assistant digital personnel (PDA). De même, de nombreuses applications permettent de jouer le contenu disponible sur un dispositif mobile directement sur la télévision grâce à des technologies telles que *Airplay*<sup>11</sup> ou *Miracast*<sup>12</sup>.

L'écran peut aussi afficher son état lorsque la télécommande est considérée comme un système. Par exemple, dans [Kobayashi *et al.*, 2009], *Rebo* est une télécommande qui héberge un avatar (Figure 3.11). L'écran de la télécommande sert à afficher le visage et les diverses émotions de l'avatar comme l'expression de la compréhension des commandes gestuelles réalisées. En complément, le corps translucide, éclairé par une lumière diffuse permet de connaître l'état de l'opération, en fonction de la couleur de l'illumination.

---

11. <http://www.apple.com/fr/airplay/>

12. <http://bit.ly/WifiMiracast>

Cette lumière visible par la vision périphérique de l'utilisateur permet de ne pas attirer totalement le regard sur la télécommande. Les auteurs catégorisent cette fonctionnalité comme un « retour périphérique ».

### 3.4 CONCLUSION

Nous avons tenté dans ce chapitre de dresser un espace de caractérisation de la télécommande qui considère le type de télécommande et l'exploitation des modalités d'interaction en entrée et en sortie.

Après avoir décrit les limitations de la télécommande traditionnelle, nous avons présenté les approches qui tentent d'améliorer cet artefact en s'inspirant des axes de classification de *Cesar et al. [2008]* et en les illustrant par des exemples de l'état de l'art. Cet espace est résumé dans le tableau 3.1 à la fin de ce chapitre.

Il reflète également plusieurs choses. D'abord, à ce jour, la recherche s'est focalisée à résoudre des problèmes particuliers comme l'accès aux commandes les plus courantes ou les plus complexes, la navigation dans des collections de contenu, la saisie de texte ou des applications particulières. Les solutions sont nombreuses et variées, mais aucune ne semblent pouvoir résoudre l'ensemble des problèmes. De nombreux auteurs estiment qu'elles sont complémentaires à la télécommande traditionnelle qui semble encore indispensable pour l'instant.

D'autre part, l'expérience télévisuelle et l'immersion au cœur des contenus est au centre de la conception des techniques. Alors que certaines permettent de focaliser l'attention visuelle sur l'écran grâce à la modalité gestuelle ou vocale, d'autres enrichissent l'expérience grâce à des écrans privés ou des retours sur la télécommande. De nouvelles approches introduisent de nouvelles formes d'exploration et de représentation des données qui accentuent l'immersion et augmentent le plaisir d'utilisation de la TV. D'autres permettent d'enrichir les programmes et complexifient encore un peu plus l'interaction avec les services.

Enfin, la plupart des travaux prennent rarement en compte le contexte d'usage de la TV différent de l'environnement de travail et de celui du bureau abordé dans le chapitre précédent.

Suite à ces observations, nous présentons dans les chapitres suivants deux techniques d'interaction gestuelle. La première est supportée par une télécom-

mande augmentée et la deuxième s'abstrait de tout dispositif physique. Lors de la conception de ces deux techniques, nous nous efforçons de garder à l'esprit le contexte de la TV interactive et les contraintes qu'il impose. Le chapitre suivant propose une classification du geste de la main. Les chapitres 5 et 6 seront destinés à la description de la conception et l'évaluation de ces techniques.

		Dimension interactionnelle						
Type de télécommande		Entrée				Sortie		
		Boutton	Geste surface	Geste main-libre	Langage naturel	Visuel	Haptique	Auditif
	Dédié	Traditionnelle Augmentée Tangible Dématérialisée	[17,28]  [9,24] [19,20,31] [40]	 [2,14,18,28,33] [7,12,33,37] [1,3,11,22,27,36,39,41]	 [5,11,38] [33] [4,13]	  [19]	 [26] [31]	  [13]
Non dédié	Accompagnement Objet du quotidien	[8,28,32,34] [6,15,23,25]	 [35]	[34] [21,30]	[34]	[8,28,32,34]	[34]	

TABLE 3.1: Catégorisation des travaux par type de télécommande et dimension interactionnelle

Liste des références présentes dans le tableau 3.1 :

1.[Aoki *et al.*, 2010], 2.[Aoki *et al.*, 2012], 3.[Bailly *et al.*, 2011], 4.[Balchandran *et al.*, 2008], 5.[Berglund et Johansson, 2004], 6.[Berglund *et al.*, 2006], 7.[Block *et al.*, 2004], 8.[Cruickshank *et al.*, 2007], 9.[Choi *et al.*, 2011], 10.[Enns et MacKenzie, 1998], 11.[Freeman T. et Craig D., 1995], 12.[Ferscha *et al.*, 2007], 13.[Goto *et al.*, 2003], 14.[Gurrin *et al.*, 2010], 15.[Hess *et al.*, 2008], 16.[Ibrahim *et al.*, 2001], 17.[Ingmarsson *et al.*, 2004], 18.[Jones *et al.*, 2010], 19.[Kobayashi *et al.*, 2009], 20.[Kimman *et al.*, 2011], 21.[Lee *et al.*, 2007], 22.[Lee *et al.*, 2009], 23.[Lima et Nevado, 2007], 24.[MacKenzie et Jusoh, 2001], 25.[Nack *et al.*, 2007], 26.[O’Modhrain et Oakley, 2003], 27.[Ren et O’Neill, 2012], 28.[Robertson *et al.*, 1996], 29.[Shoemaker *et al.*, 2010], 30.[Schmidt *et al.*, 2004], 31.[Tahir *et al.*, 2008], 32.[Tsekleves *et al.*, 2007], 33.[Tseng et Kolko, 2005], 34.[Turunen *et al.*, 2009], 35.[Vatavu et Pentiuc, 2008], 36.[Vatavu, 2012], 37.[Villar et Hodges, 2010], 38.[Wittenburg *et al.*, 2006], 39.[Yoo *et al.*, 2011], 40.[Zhang *et al.*, 2011], 41.[Zigelbaum *et al.*, 2010]



---

## ESPACE DE CARACTÉRISATION DU GESTE DE LA MAIN POUR LA TÉLÉVISION INTERACTIVE

---

L'étude de l'interaction gestuelle est un sujet qui débute dans le domaine du langage et de la communication. De nombreux auteurs se sont intéressés à la gesticulation où le geste est un support de la parole [Efron, 1941; Kendon, 1980; McNeill, 1992]. Ces travaux ont inspiré la communauté en interaction-homme machine pour guider la conception de gestes permettant de communiquer avec les systèmes.

### 4.1 TAXONOMIES EXISTANTES EN INTERACTION HOMME-MACHINE

Wexelblat [1994] propose une caractérisation pour un analyseur de geste dans le contexte des interfaces multimodales. Son but est de concevoir un système de reconnaissance capable de comprendre les gestes naturels des utilisateurs sans avoir à leur imposer des configurations particulières ou des gestes prédéfinis. Cette caractérisation tient compte en particulier de la sémantique communicative des gestes. Cette caractérisation compte quatre classes et neuf types de geste : les gestes *symboliques* / *modaux*, *pantomimiques*, *iconiques*, *déictiques*, de « Lakoff », et les gestes dit de « battements » / de « Butterworths » ou *autocentrés*.

Alors que les gestes *symboliques* sont des gestes qui sont directement liés à un sens comme lever le pouce pour signaler que tout va bien (Figure 4.1), les geste *modaux* sont ceux qui changent le sens de la communication et qui indiquent le mode d'interprétation que devrait adopter un interlocuteur. Par exemple, une personne qui n'est pas sûre de ses propos pourra hausser les épaules. Les gestes *pantomimiques* représentent les gestes effectués par une personne pour mimer l'usage et l'action sur des objets. Les gestes *iconiques* représentent les gestes de la main qui incarnent un objet dans une scène ou un événement. Par exemple, pour décrire un objet qui tombe, la main pourra réaliser un geste de haut en bas. Les gestes *déictiques* sont les gestes qui pointent





FIGURE 4.1: Un geste symbolique avec les pouces levés vers le haut pour signaler que tout va bien. Photo de Andrew Beresford sur Flickr.

vers un objet, un groupe d'objet, un endroit ou une direction. Les gestes de *Lakoff* accompagnent une métaphore verbale spatialisée et montre sa direction. Un exemple, en anglais, est l'expression « *I'm feeling down* » signifiant une mauvaise humeur qui est souvent accompagnée d'un geste vers le bas. Les battements marquent le rythme et la vitesse de discours. Les gestes dit de *battement* permettent à un orateur de montrer quels mots sont importants dans le discours et d'aider son audience à rester attentive. Les gestes dit « *Butterworth* » représentent les gestes durant une conversation non liés au discours. Enfin, les gestes *autocentrés* sont les gestes qui indiquent l'état mental d'une personne comme sa nervosité ou sa détermination.

Inspiré par l'interaction entre un musicien et son instrument, [Cadoz \[1994\]](#) définit la communication instrumentale. Pour cet auteur, le canal gestuel, incarné par la main, possède trois fonctions : une fonction *épistémique*, *ergotique* et *sémiotique*. La fonction *épistémique* du geste décrit un geste exploratoire qui permet d'acquérir de l'information grâce aux différents récepteurs de la main. La fonction *ergotique* exprime la capacité de la main à transmettre de l'énergie aux objets qu'elle manipule. Enfin, les gestes *sémiotiques* produisent un message informationnel.

En combinant ces trois fonctionnalités, il propose le geste instrumental comme une modalité de communication qu'il décrit avec trois nouvelles caractéristiques : le geste d'*excitation*, le geste de *modulation* et le geste *déictique* (ou de sélection) (Figure 4.2). Pour comprendre cette typologie, il est nécessaire de considérer un instrument dont la fonction première est de convertir le geste en un phénomène sensible (acoustique, visuel...). Le geste d'*excitation* est la composante du geste qui fournit de l'énergie au phénomène sensible.



FIGURE 4.2: Une violoniste. Le geste d'excitation est illustré par la manipulation de l'archet sur les cordes, le geste de modulation par l'accord des doigts sur les cordes. Photo de Ben Clement sur FlickrR.

Le geste de *modulation* apporte une nouvelle dimension porteuse de variations expressives à l'information délivrée par l'instrument. Ces modulations peuvent être discrètes ou continues. Enfin, le geste *déictique* a pour rôle de sélectionner un instrument élémentaire. A la différence du geste d'*excitation* ou de *modulation*, ce geste ne communique pas d'énergie ou ne modifie pas l'objet.

Hummels et Stappers s'appuient sur la définition suivante du geste : « *un geste est un mouvement du corps d'une personne qui transmet une information porteuse de sens à soi-même ou un partenaire en communication* » pour concevoir un espace de caractérisation de geste pour le design industriel [Hummels et Stappers, 1998]. Insatisfaits des classifications basées sur la gesticulation qui ne tiennent compte que des aspects spatiaux, temporels et symboliques, ils proposent une classification qui tient compte du contexte des concepteurs, de la spécificité de la tâche et des émotions. Cette classification comprend les gestes *spatiaux* qui se réfèrent à l'interaction avec un objet physique en trois dimensions. Par exemple, ces gestes peuvent être : utiliser un objet en le saisissant, indiquer un ou plusieurs objets en le pointant, décrire la forme d'un objet comme en l'ébauchant sur une surface ou encore manipuler l'objet en le contractant ou le tirant pour en changer la forme. Elle comprend également les gestes *passifs*. Ces gestes supportent le processus de conception mais n'interviennent pas dans la manipulation de l'objet. Enfin, les gestes *symboliques* représentent un concept culturel comme mettre ses doigts en forme de "V" pour représenter la victoire. Parallèlement aux différents types de geste, la dimension affective, basée sur un modèle psychologique de l'émotion, décrit l'expression émotionnelle avec deux axes matérialisant le niveau d'activation

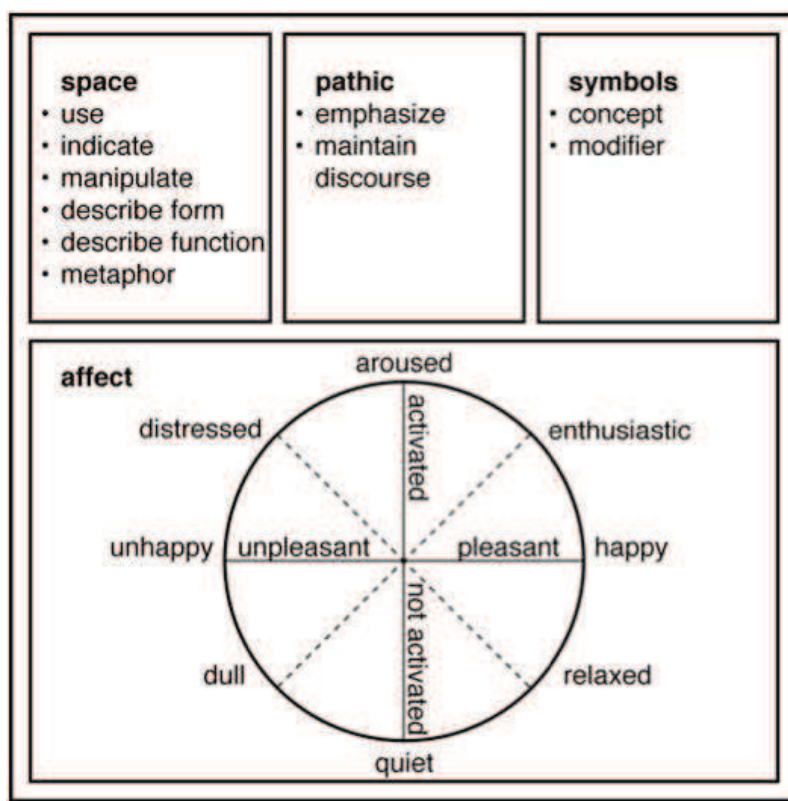


FIGURE 4.3: Une proposition de caractérisation du geste dans le contexte du design industriel [Hummels et Stappers, 1998].

et le niveau d'agrément (Figure 4.3).

Quek *et al.* [2002] proposent de distinguer trois types de gestes : les gestes *manipulateurs*, les gestes *sémaphoriques* et les gestes liés à la *conversation*. De leur perspective, les gestes *manipulateurs* sont les gestes consacrés au contrôle d'une entité par l'application d'une relation étroite entre le mouvement du geste de la main et du bras, et l'entité manipulée. Les gestes *sémaphoriques* sont inspirés des sémaphores, un canal de communication visuel utilisant des drapeaux à bout de bras pour transmettre des caractères. Ils représentent les gestes utilisés par les systèmes qui emploient un dictionnaire de gestes statiques (immobiles dans le temps comme les postures) ou dynamiques (mobiles dans le temps) des bras ou des mains. Par exemple, *Finger Count* [Bailly *et al.*, 2010] permet de sélectionner une commande dans un menu en montrant une combinaison de doigts avec les deux mains au système (Figure 3.19).

Basée sur les classifications précédentes [Quek *et al.*, 2002; Wexelblat, 1994], Karam *et m. c. schraefel* [2005] présentent une taxonomie exhaustive qui dé-

finit cinq types de geste : les gestes *déictiques*, *manipulateurs*, *sémaphoriques*, *gesticulatoires* et les *langages gestuels*. Un peu à la manière de la définition donnée par Cadoz [1994], les gestes *déictiques* ont une fonction de localisation et de sélection spatiale. Les auteurs apportent une granularité intéressante quant à la description des gestes *manipulateurs*. Leur description prend en compte les degrés de liberté disponibles en entrée et la correspondance avec le nombre de degrés contrôlables de l'objet d'interaction (Figure 4.4). Les auteurs affinent la notion de geste *sémaphorique* avec l'introduction des gestes directionnels rendus populaires grâce aux *Marking Menus* [Kurtenbach et al., 1993]. Enfin, ils distinguent les gestes issus de la gesticulation et les langages basés sur les gestes qui permettent une communication continue, comme par exemple le langage des signes.

En analysant des gestes récoltés dans le cadre d'une étude de définition de gestes par les utilisateurs (Figure 4.5), Wobbrock et al. [2009] proposent une taxonomie pour l'interaction sur surface interactive. Quatre caractéristiques de gestes sont proposés : la *forme* des gestes, la *nature* des gestes, la *relation* du geste et le *flux* du geste. La *forme* du geste tient compte du contact du doigt ou de la main avec la surface et de la dynamique du geste. La *nature* du geste permet d'établir une relation entre le geste et son effet et décrit l'aspect sémantique du geste avec quatre variables : les gestes *symboliques*, *physiques*, *métaphoriques* et *abstraites*. Les gestes *symboliques* correspondent à une description visuelle comme tracer un accent circonflexe pour réaliser l'action d'insérer un objet. Les gestes *physiques* ont le même effet sur une table avec des objets physiques. Les gestes *métaphoriques* sont les gestes qui agissent sur ou/et avec un objet comme si on utilisait un autre objet. Par exemple, faire un geste de balayage avec le doigt sur la surface d'un écran pour tourner la page d'un document numérique. Enfin, les gestes *abstraites* sont définis comme les gestes qui n'ont aucune relation symbolique, physique ou métaphorique avec l'action à laquelle ils sont associés.

Les auteurs soulignent l'importance du contexte pour définir le geste et définissent une dimension de relation entre le geste et l'objet. Ils distinguent d'abord les gestes *centrés sur l'objet d'interaction* comme le geste de pincement d'un objet virtuel pour réduire sa taille. Les gestes *dépendants du monde* sont définis vis à vis du monde comme appuyer sur un coin de l'écran ou encore déplacer un objet à l'extérieur de l'écran. Les gestes *indépendants du monde* ne dépendent ni de l'objet d'interaction ni du monde dans lequel ils évoluent. Parfois, les dépendances peuvent être mixtes dans le cas où l'opérateur utilise ses deux mains simultanément. Alors qu'une main manipule un objet, l'autre

main agit ailleurs sur l'écran.

Enfin, le flux du geste, *continu* ou *discret*, correspond au type de reconnaissance nécessaire pour reconnaître le geste, un flux continu nécessitant une reconnaissance permanente.



FIGURE 4.4: Un geste manipulateur : l'utilisateur manipule un cube physique pour explorer l'interface affichée sur l'écran [Block *et al.*, 2004].

Freeman *et al.* [2009] étendent la taxonomie de Wobbrock *et al.* [2009] en ajoutant trois nouvelles dimensions à la catégorie forme : la pose de *détection*, la pose de *continuation* et le *mouvement*. Cette nouvelle caractérisation fait la différence entre les gestes réalisés du bout des doigts et les autres parties du doigt ou de la main. La pose de *détection* correspond au début du geste et peut s'exprimer par quatre types de poses différents : le doigt unique, le multi-doigts, la forme simple, ou les formes multiples correspondant à l'interaction bi-manuelle. Les auteurs définissent la pose de *continuation* comme tout changement de pose se déroulant après la pose de *détection*. Cette phase peut être *statique* si la posture de la main reste la même après la détection ou *dynamique* si la posture de la main change, avec par exemple le contact de doigts avec la surface à posteriori. Enfin, le mouvement est défini par la présence ou non d'un chemin dans le geste. Notons que le geste des doigts lorsque la main est immobile n'est pas considéré comme un mouvement mais plutôt comme une pose de continuation dynamique.

Dans leur espace de caractérisation des interactions gestuelles physiques sur dispositif mobile, Baglioni *et al.* [2009] proposent quatre types de gestes : les gestes *physiques*, les gestes *métaphoriques*, les gestes *symboliques* et les gestes *abstraites*. Les gestes *physiques* correspondent aux gestes ergotiques de Cadoz



[1994] et aux gestes manipulateurs de [Quek et al. \[2002\]](#). Les gestes *métaphoriques* associent deux champs d'application différents. Ils se fondent sur le concept de transposition et désignent la réutilisation d'un geste appartenant à un champ d'application particulier pour réaliser une tâche dans un autre champ d'application. Ce sont également des gestes pantomimiques au sens de [Wexelblat \[1994\]](#). La catégorie des gestes *symboliques* est comparable à la catégorie des gestes sémaphoriques de [Karam et m. c. schraefel \[2005\]](#) à la différence près qu'ils portent une information sémantique associée à une commande. Enfin la catégorie des gestes abstraits est similaire à la classification de [Wobbrock et al. \[2009\]](#).

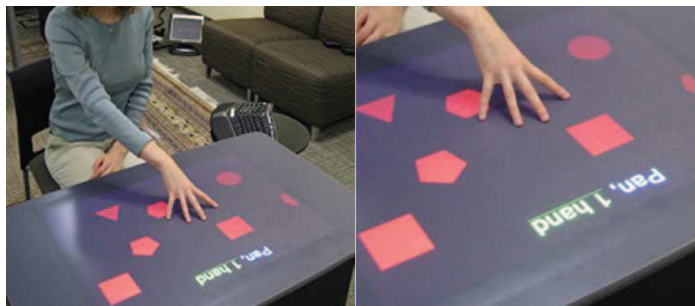


FIGURE 4.5: Un utilisateur réalise un geste de défilement à la suite d'une démonstration de l'effet du geste [[Wobbrock et al., 2009](#)].

Tous ces gestes peuvent être réalisés au moyen de deux types de mouvement : les mouvements *fluides* et les mouvements *impulsifs*. Les auteurs définissent le mouvement *fluide* comme un mouvement dont la signature accélérométrique présente des amplitudes de variations faibles et étalées sur des durées longues. Les mouvements *impulsifs*, quant à eux, sont caractéristiques d'une signature qui présente des amplitudes élevées sur des durées très courtes. Cette classification intègre également la notion de contrôle. Ces gestes permettent de manipuler les variables d'une application de deux manières différentes : d'une manière *continue* ou *discrète*. Le contrôle continu en général requiert une certaine précision que seuls les mouvements fluides sont capables d'offrir. Le contrôle discret souvent associé aux gestes impulsifs permet plutôt d'exécuter des commandes.

De la même manière que [Wobbrock et al. \[2009\]](#), [Ruiz et al. \[2011\]](#) proposent une taxonomie pour les gestes de mouvement pour l'interaction mobile basée sur l'analyse d'un jeu de gestes définis par les utilisateurs. Il classifie les gestes selon quatre caractéristiques : la *correspondance*, les caractéristiques *physiques*, les dimensions de *mouvement* et la *complexité* du geste. De nombreuses notions sont inspirées des travaux de [Wobbrock et al. \[2009\]](#) telle que

la *nature* du geste. La dimension du *contexte* décrit si le geste nécessite un contexte spécifique comme rapprocher le téléphone de son oreille pour décrocher un appel entrant ou s'il peut être réalisé indépendamment comme secouer son téléphone pour revenir à l'écran d'accueil. Les caractéristiques *physiques* font référence à l'impulsion du geste déjà abordée dans [Baglioni *et al.*, 2009] avec trois niveaux différents et la combinaison des différentes dimensions qui permettent de décrire le mouvement. On retrouve aussi, dans la dimension temporelle de cette classification, la notion de contrôle de Baglioni *et al.* [2009] qui décrit si l'action se réalise pendant le geste ou après la fin du geste. Enfin, la *complexité* du geste est soulignée en distinguant les gestes *simples* et les gestes *composés*.

Scoditti *et al.* [2011] proposent une nouvelle taxonomie basée sur les données capturées uniquement par des accéléromètres. Cette taxonomie, centrée sur une perspective de l'utilisateur plutôt que du dispositif, propose de couvrir une dimension *sémantique*, *syntactique*, *lexicale* et *pragmatique*. La dimension *sémantique* correspond aux tâches d'interaction réalisées par l'utilisateur. Ces tâches sont inspirées des travaux de Foley *et al.* [1984]. Les auteurs caractérisent la dimension *syntactique* d'une technique d'interaction par deux variables appelées modificateurs syntaxiques. Ces modificateurs représentent l'existence ou l'absence de délimiteur d'interaction et le type de contrôle associé, contrôle de position, de vitesse ou d'accélération. La dimension *lexicale* représente le lexique du dispositif d'entrée correspondant aux actions physiques transmises par l'utilisateur. Ces actions physiques sont divisées en deux groupes : les actions conscientes qui sont effectuées explicitement par l'utilisateur et les actions inconscientes que le système réalise au nom de l'utilisateur. Par exemple, pour économiser de l'énergie, le téléphone peut interrompre l'affichage de l'écran lorsque l'utilisateur approche son téléphone de l'oreille pendant un appel téléphonique. Cette dimension permet également de distinguer les interactions qui ont lieu au premier plan (explicites) avec les transformations géométriques impliquées dans les mouvements de l'utilisateur et à l'arrière plan (implicites). Enfin, la dimension *pragmatique* exprime la distance entre les attentes de l'utilisateur et la sémantique de la technique d'interaction. Cette dimension, également décrite par le *degré d'intégration* dans le modèle de l'interaction instrumentale de Beaudouin-Lafon [2000], s'intéresse à la correspondance entre les dimensions en entrée et en sortie de l'interaction. Par exemple, pour visualiser une carte dans [Rekimoto, 1996], incliner le dispositif dans une direction permet de déplacer la carte dans cette même direction, créant ainsi une correspondance directe. Dans *TiltType*, pour saisir du texte, incliner le téléphone mobile dans une direction permet de sélectionner un caractère créant alors une correspondance indirecte [Wigdor et

Balakrishnan, 2003].

*BodyScape* est un espace de conception qui permet de décrire, classer et comparer des techniques d'interaction multi-surfaces [Wagner *et al.*, 2013]. Plus précisément, cet espace s'appuie sur la position relative des dispositifs d'entrée et de sortie à celle de l'utilisateur et l'implication des parties du corps dans l'interaction pour décrire les techniques. Dans cet espace, la position de chaque dispositif d'entrée ou de sortie peut être soit relative au corps ou soit fixée dans le monde. Par exemple, les techniques d'interaction gestuelle sur le corps sont considérées comme relatives au corps, tout comme l'affichage sur un écran intégré à un dispositif tenu par l'utilisateur. Une technique d'interaction réalisée sur une table interactive sera plutôt considérée comme une technique fixe dans le monde. Il en va de même, par exemple, pour un affichage distant sur une télévision. Ces propriétés décrivent comment une technique peut restreindre la position et l'orientation du corps de l'utilisateur pendant l'interaction et impacter ses performances. C'est d'ailleurs pourquoi les auteurs différencient les techniques d'interaction gestuelle contraintes par une surface et celles qui sont mains-libres. Ils considèrent également le nombre de parties du corps impliquées dans l'interaction. L'interaction des membres du corps sur d'autres membres peuvent affecter leur disponibilité et restreindre ainsi l'implication du corps dans l'interaction. Enfin, notons que les auteurs distinguent deux types de combinaisons de techniques d'interaction pour les tâches complexes. La combinaison est qualifiée de série lorsque deux techniques sont réalisées l'une après l'autre. Elle est dite parallèle lorsque les deux techniques sont réalisées en même temps.

Ces différentes classifications couvrent une grande partie de l'interaction gestuelle en interaction Homme-Machine. De nombreux concepts similaires se recouvrent alors que certains concepts sont intrinsèques aux différents type d'interaction. Dans la suite, nous proposons d'unifier ces concepts à travers l'exploration de la littérature discutée auparavant. Nous présentons un espace de classification qui distingue le type d'interaction et décrit le geste par sa fonction et ses différentes étapes d'exécution.

#### 4.2 ESPACE DE CARACTÉRISATION PROPOSÉ

Cet espace de classification s'inspire des travaux de [Wexelblat, 1994; Cadoz, 1994; Quek *et al.*, 2002; Karam *et m. c. schraefel*, 2005; Wobbrock *et al.*, 2009; Freeman *et al.*, 2009; Baglioni *et al.*, 2009; Ruiz *et al.*, 2011; Scoditti *et al.*, 2011]. Il présente le geste en fonction du type d'interaction considéré, de sa



fonction, de son anatomie, de sa combinatoire et du contexte dans lequel il est réalisé (Tableau [4.1](#)).

Espace de caractérisation du geste de la main				
Type d'interaction	Surface	Non instrumenté		
		Instrumenté <i>Interaction avec un stylet sur une tablette par exemple</i>		
	Mains libres	Non instrumenté		
		Instrumenté <i>Interaction avec un téléphone mobile ou une télécommande par exemple</i>		
Fonction	Épistémique		<i>Permet d'acquérir de l'information</i>	
	Ergotique	Physique		<i>Manipule un objet</i>
		Métaphorique		<i>Emprunte des techniques de manipulation pour d'autres objets</i>
	Sémiotique	Sémaphorique		<i>Communique via des symboles avec l'environnement</i>
		Déictique		<i>Indique l'identité ou la location d'un élément particulier</i>
		Abstrait		<i>Ne possède aucun lien sémantique avec la commande associée</i>
Anatomie	Délimiteur	Aucun		<i>Ne modifie pas l'action du geste</i> <i>Modifie l'action du geste</i>
		Statique		
		Dynamique		
	Cinématique	Trajectoire	Aucune	<i>Rectiligne, courbe ou elliptique</i> <i>Symboles complexes porteurs de sens</i>
			Élémentaire	
			Complexe	
Combinaison	Orientation			
	Accélération			
	Vitesse			
Contexte	Mécanique			
	Perceptif			
	Social			

TABLE 4.1: Espace de caractérisation du geste de la main pour la télévision interactive

#### 4.2.1 *Type*

En interaction homme-machine, à notre connaissance, la littérature fait état de deux types d'interaction gestuelle : l'interaction gestuelle dans lequel le geste est libre ou bien le geste est contraint par une surface. Ces deux types d'interaction peuvent être instrumentés. Dans le cas du geste libre, l'interaction peut être par exemple réalisée à l'aide d'une télécommande ou d'un téléphone mobile. Dans le cas où le geste est contraint par une surface, il peut être effectué à l'aide d'un stylo numérique [Malacria et Lecolinet, 2008], d'un écran tactile, d'une tablette graphique, etc.

Le geste libre correspond à la forme du geste la moins restrictive. Il est seulement contraint par les limites physiologiques du mouvement des parties du corps dans l'espace. Le geste contraint par une surface correspond à l'interaction dite tactile. Dans ce contexte, le geste est supporté par une surface qui offre un certain confort d'utilisation et un espace de guidage grâce au retour haptique passif résultant du contact du doigt avec cette surface. Finalement, le geste instrumenté correspond au geste effectué à l'aide d'un instrument. Dans cette configuration, la ou les mains manipule(nt) un instrument, généralement un dispositif mobile équipé de capteurs d'orientation, qui permet de digitaliser les mouvements du geste dans l'air. Les gestes réalisés dans cette condition peuvent dépendre de certains facteurs tels que la présence d'un écran embarqué comme dans le cas des téléphone mobiles, qu'il est en général nécessaire de voir pour accomplir l'action. Dans d'autres cas, cet instrument peut aussi jouer le rôle du prolongement du bras à la manière d'une baguette magique [Wilson et Shafer, 2003].

Ces deux types de gestes partagent des propriétés susceptibles de fournir une catégorisation générale du geste de la main. Nous verrons que, en fonction du type de geste, ces propriétés s'expriment différemment. Elles déterminent la fonction du geste et son anatomie.

#### 4.2.2 *Fonction*

Pour décrire les fonctions du geste nous reprenons la définition de Cadoz [1994] qui définit le canal gestuel par trois fonctions : une fonction *épistémique*, *ergotique* ou *sémiotique*.

La fonction *épistémique* du geste décrit un geste exploratoire permettant d'acquérir de l'information grâce aux différents récepteurs de la main (Figure 4.6). Ce geste fait intervenir le sens tactilo-kinesthésique ou haptique et

le sens proprioceptif. Le sens haptique est le résultat de la stimulation des récepteurs de la peau, des muscles des tendons et des articulations générée par l'exploration manuelle avec contact d'un objet dans l'espace [Gentaz *et al.*, 2008]. Le sens proprioceptif, quant à lui, interprète les signaux du mouvement du corps. Plus précisément, cela inclut la perception de la position et le mouvement des membres du corps, la perception de la tension et la force, la perception de l'effort et la perception de l'équilibre [Proske et Gandevia, 2012]. Cette fonction peut être illustrée par le geste résultant du déplacement de la main pour apprécier la texture, la température, la rigidité ou toute autre propriété d'une surface.

D'ailleurs, de nombreux travaux s'intéressent à tirer avantage de cette capacité. Par exemple, certains systèmes comme *TeslaTouch* [Bau *et al.*, 2010] et *REVEL* [Bau et Poupyrev, 2012] utilisent le principe de l'électrovibration pour altérer virtuellement la texture d'une surface. Il est alors possible de percevoir le relief d'une image de la main ou encore d'altérer la perception de la texture d'un objet. D'une autre manière, la technologie *STIMTAC* [Biet *et al.*, 2007] permet d'altérer le coefficient de friction d'une surface tactile comme dans *Surfpad* pour faciliter le pointage [Casiez *et al.*, 2011]. Le sens proprioceptif peut aussi jouer un rôle dans l'exploration gestuelle. Roudaut *et al.* [2013] utilisent le mouvement de la surface sous le doigt pour transmettre des symboles sous forme de geste 2D. Le mouvement de la surface entraîne un mouvement passif du doigt le long d'un chemin. Une autre voie qu'explore Follmer *et al.* [2012] est la capacité à distinguer la déformation d'un objet. Grâce à un système pneumatique, les auteurs altèrent la forme d'un objet dynamiquement.

La fonction *ergotique* exprime la capacité de la main à transmettre de l'énergie aux objets qu'elle manipule. Elle rejoint la définition des gestes manipulateurs dans [Quek *et al.*, 2002; Karam et m. c. schraefel, 2005] comme des gestes dévolus au contrôle d'entité par l'application d'une relation étroite entre le mouvement du geste de la main et du bras, et l'entité manipulée. Ils représentent également l'axe lexical de la classification de Scoditti *et al.* [2011] qui décrivent les actions physiques appliquées au dispositif. Cette fonction compte deux types de gestes qui sont distingués dans [Wobbrock *et al.*, 2009; Baglioni *et al.*, 2009; Ruiz *et al.*, 2011] : les gestes *physiques* et les gestes *métaphoriques*. Les gestes *physiques* sont un transfert direct des gestes du monde physique dans le monde numérique à l'image de la manipulation directe. Par exemple, déplacer une photographie posée sur une table du doigt dans le monde réel se réalise de la même manière sur une surface interactive. Les gestes *métaphoriques* sont des actions sur un objet qui empruntent des tech-



FIGURE 4.6: Fonction épistémique : l'utilisateur explore des textures tactiles virtuelles sur un objet réel [Bau et Poupyrev, 2012].

niques de manipulation utilisées pour d'autres objets. Un exemple est l'action d'éloigner son téléphone portable de son visage pour diminuer le niveau d'agrandissement de l'affichage de l'écran comme on manipulerait une loupe [Ruiz *et al.*, 2011]. Ces gestes sont aussi appelés gestes pantomimiques [Wexelblat, 1994].

La fonction *sémiotique* désigne la capacité des messages à fournir des informations destinées à l'environnement. Cette catégorie comprend les gestes *sémaphoriques* de [Quek *et al.*, 2002] qui sont communicatifs par le fait qu'ils communiquent via un univers de symbole à la machine. Ils sont généralement utilisés pour communiquer à distance avec les systèmes [Karam et m. c. schraefel, 2005]. Aussi appelés gestes symboliques [Wexelblat, 1994; Hummels et Stappers, 1998; Wobbrock *et al.*, 2009; Ruiz *et al.*, 2011], ils sont une description visuelle ayant une signification particulière et permettent d'invoquer une commande. Cette représentation peut prendre la forme d'une posture statique de la main ou d'un tracé dessinant un symbole [Wobbrock *et al.*, 2009]. Les gestes de nature *déictique* forment la catégorie des gestes qui permettent de sélectionner un élément, un groupe d'éléments ou un emplacement [Wexelblat, 1994; Karam et m. c. schraefel, 2005]. Ils sont différents des gestes symboliques par le fait qu'ils servent à cibler ou indiquer l'identité d'un élément particulier. Enfin les gestes *abstraites* sont les gestes qui n'ont aucun lien sémantique avec la commande ou la tâche à laquelle ils sont associés. Par exemple, les gestes directionnels popularisés grâce aux *Marking Menus* [Kurtenbach *et al.*, 1993] font partie de cette catégorie. Notons que ces gestes directionnels correspondent aussi aux gestes abstraits dans [Wobbrock *et al.*,

2009; Ruiz *et al.*, 2011].

#### 4.2.3 Anatomie gestuelle

Dans le monde réel, le geste est un flux continu du mouvement d'une ou de plusieurs parties du corps. En identifiant des structures temporelles dans la réalisation des gestes, Wilson *et al.* [1996] décrivent le geste comme une succession de phases dépendantes du temps. Chaque geste est délimité par une phase où les mains sont dans une situation de repos. Ils introduisent les gestes tri-phasiques qui se composent d'une séquence d'un mouvement de transition de la phase de repos à l'espace du geste, puis d'une posture et d'un retour à la phase de repos. Baudel et Beaudouin-Lafon [1993] définissent des zones actives délimitant spatialement les gestes et, une combinaison de positions de départ et d'arrivée pour les différencier entre eux. Ces approches sont intéressantes car elles proposent de segmenter le geste et permet, d'une part, de séparer les gestes d'intérêt des gestes quelconques (ou de repos) et, d'autre part, identifier les différentes phases du geste pour en améliorer la reconnaissance. D'autres catégorisations du geste suivent cette approche. Par exemple, [Freeman *et al.*, 2009] proposent une pose de détection et de continuation correspondant au délimiteur gestuel, et un mouvement constitué d'un chemin pour décrire le geste dans leur taxonomie des gestes de la main sur surface interactive. Un autre exemple, dans le cadre des interactions basées sur des accéléromètres, est les modificateurs syntaxiques de la classification de [Scoditti *et al.*, 2011] qui possèdent une variable indiquant la présence d'un délimiteur gestuel ou non. Leur axe lexical qui désigne les actions appliquées au dispositif est décrit, entre autres, avec les transformations affines impliquées dans le mouvement du dispositif.

La littérature sur les techniques d'interaction gestuelle fait état de la nécessité de délimiter les gestes pour distinguer les gestes intentionnels des gestes ou des mouvements du quotidien [Wilson *et al.*, 1996; Ruiz et Li, 2011]. Une autre partie de la littérature se concentre sur la description du mouvement réalisé pendant ces gestes pour inférer les actions sur les objets d'interaction. Le geste peut être complexe : il peut être une séquence ou une combinaison de gestes élémentaires. En suivant ce schéma, nous présentons ici ces caractéristiques qui permettent de décrire un geste.

#### 4.2.3.1 *Délimiteurs*

Nous avons vu précédemment que le geste doit être délimité pour extraire les informations pertinentes à la réalisation de l'action. Le rôle de ces délimiteurs est de différencier les gestes d'intérêt des gestes quelconques en indiquant au système le moment auquel il doit écouter. Les travaux antérieurs de caractérisation du geste font état de deux types de délimiteurs : les délimiteurs statiques et les délimiteurs dynamiques.

Les *délimiteurs statiques* correspondent aux délimiteurs qui ne modifient pas l'action déclenchée par le geste. Ils sont en général des postures facilement reconnaissables par le système. Pour l'interaction sur surface tactile ce délimiteur est souvent représenté par un contact d'une partie du corps avec la surface. Ces délimiteurs peuvent aussi porter d'autres informations que le début du geste. Dans [Freeman *et al.*, 2009], les auteurs rapportent quatre délimiteurs gestuels statiques différents pour l'interaction avec une surface interactive. En plus du contact d'un seul doigt avec la surface, les auteurs enrichissent ce vocabulaire par le contact simultané de plusieurs doigts, le contact avec la main entière ou le contact avec deux mains. Dans un autre registre, l'interaction pour dispositif mobile nécessite également de délimiteurs pour capturer les informations de mouvement pertinentes. Dans *TiltText*, Wigdor et Balakrishnan [2003] proposent de délimiter le geste d'inclinaison du téléphone par l'appui d'une touche du clavier. Tant que la touche du clavier est enfoncée, le système enregistre les données d'inclinaison. Lors du relâchement de la touche, le système interprète le geste.

Une autre approche est l'utilisation d'un geste à part entière comme délimiteur. *DoubleFlip* est un geste auto-délimité : il se démarque facilement des autres gestes disponibles pour manipuler un smartphone (Figure 4.9) et, de plus, il est robuste à la reconnaissance fortuite de tout autre geste non affecté au contrôle du dispositif [Ruiz et Li, 2011]. Il est alors possible d'utiliser ce geste avant ou/et après pour délimiter un autre geste. Les *JerkTilts* dont la signature cinématique est unique présentent des propriétés similaires [Baglioni *et al.*, 2011]. L'interaction main libre utilise généralement une posture de la main comme délimiteur. Par exemple dans [Freeman T. et Craig D., 1995; Bailly *et al.*, 2011], le système commence à interpréter les gestes lorsque l'utilisateur montre la paume de sa main à l'écran distant. Pour sortir du mode gestuel, il lui suffit de fermer la main.

Les *délimiteurs dynamiques* correspondent aux délimiteurs qui peuvent modifier une action déclenchée par le geste. En général, ils correspondent à un changement de posture précédant le mouvement du geste. Appelés poses de

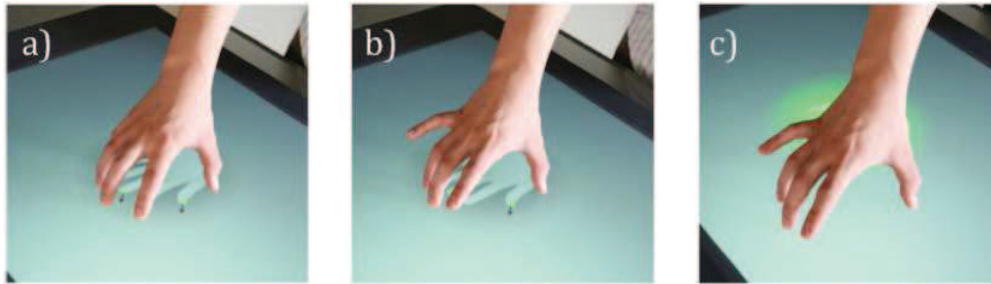


FIGURE 4.7: Délimiteur dynamique : a - L'utilisateur pose les 5 doigts sur la surface pour calibration. b - Il retire les doigts non utilisés pour l'accord. c - Il réalise ensuite un mouvement directionnel [Lepinski *et al.*, 2010].

continuation dans [Freeman *et al.*, 2009], les auteurs définissent ces gestes comme une modification de la posture de la main qui n'entraîne pas un changement d'emplacement. *Finger Count* [Bailly *et al.*, 2010] illustre cette définition. Pour faire apparaître un menu sur une surface interactive, il suffit de poser ses doigts sur la surface. Le nombre de doigts conditionne l'élément du menu qui est sélectionné. En pianotant ses doigts sur la surface, il est possible de naviguer dans le menu. La sélection est effective lorsque le contact de tous les doigts et de la surface est rompu. Dans ces exemples, la main n'a pas changé d'emplacement. Un autre exemple est la technique de *Lift-and-Stroke* dans [Lepinski *et al.*, 2010] qui impose à l'utilisateur à poser les cinq doigts de la main pour calibrer le système avant de lever les doigts nécessaires pour sélectionner un accord (Figure 4.7).

#### 4.2.3.2 Cinématique

La cinématique est l'étude des mouvements d'un corps. Pour décrire ce mouvement, cette branche de la mécanique s'appuie sur la trajectoire des corps et leurs propriétés différentielles comme la vitesse ou l'accélération. Nous considérons le mouvement du geste de la main comme le mouvement d'un système mécanique. Il correspond au mouvement effectué entre le moment de l'activation d'un délimiteur gestuel et sa désactivation.

La *trajectoire* du geste est la ligne décrite par le mouvement du geste. Lorsqu'elle existe elle peut prendre différentes formes orientées élémentaires, comme des traits directionnels, des formes géométriques simples sans signification particulière ou des symboles plus complexes porteurs de sens comme vu précédemment. Ce mouvement peut également être caractérisé par son accélération et sa vitesse.



Il n'y a aucune trajectoire quand un geste ne se déplace pas au cours du temps. Ce sont les gestes posturaux vus précédemment. La forme de la trajectoire peut être élémentaire. Par exemple, dans le cadre de l'interaction sur surface interactive, les *Marking Menus* se basent sur l'orientation de gestes rectilignes pour choisir un élément dans un menu circulaire [Kurtenbach et Buxton, 1991]. Une autre propriété de l'espace de conception des *Marking Menus* est la longueur du geste [Zhao et al., 2006]. On peut également retrouver des gestes avec une trajectoire élémentaire dans les gestes de glissements (*Flick*) qui sont très répandus sur les surfaces tactiles. Il est aussi possible de retrouver ces gestes pour l'interaction main libre où ils permettent de naviguer dans un menu linéaire dans [Lee et al., 2009] et de sélectionner une commande à la manière des *Marking Menus* dans [Bailly et al., 2011].

Certains gestes directionnels présentent un rayon de courbure. Les *Flower Menus* en donnent un exemple [Bailly et al., 2008]. Les *Flower Menus* proposent de compléter les gestes rectilignes par des gestes avec une trajectoire courbée permettant ainsi d'augmenter la capacité d'adressage de commandes. Suivant les paramètres de cette courbe, il est possible de sélectionner une commande différente. Par ailleurs, la structure des dispositifs peut contraindre l'utilisateur à effectuer des gestes courbes. Dans [Voelker et al., 2012] par exemple, les auteurs étudient les gestes de glissement sur une surface interactive composée d'une partie horizontale et d'une partie verticale raccordées par une partie courbée.

Les gestes elliptiques sont des gestes de forme courbe avec une trajectoire fermée. Dans [Malacria et al., 2010], *CycloZoom* est une technique d'interaction qui propose d'utiliser des gestes de forme elliptique pour contrôler le niveau de zoom d'une application sur surface tactile. Les gestes de cette forme sont également réalisables pour de l'interaction gestuelle main libre. Cependant, ils semblent moins performants dans ce contexte [Nancel et al., 2011]. Notons au passage que les gestes circulaires sont un cas particulier des gestes elliptiques. Les gestes de forme complexe peuvent être décomposés grâce aux gestes de forme élémentaires, en particulier les gestes rectilignes et courbes. Néanmoins, l'information essentielle transmise par ce type de geste est la forme globale du geste. Généralement, cette forme est la représentation gestuelle d'un symbole comme dans [Enns et MacKenzie, 1998], où effectuer des gestes sur le trackpad d'une télécommande en forme de chiffre permet de changer de chaîne. Un autre exemple est *Gesture Avatar* [Lü et Li, 2011] qui permet l'association d'un geste symbolique à un widget sur dispositif mobile pour en prendre le contrôle. Dans cet exemple, l'utilisateur associe un symbole au curseur d'une barre de défilement. La taille du symbole permet

de créer un objet manipulable qu'il est possible de déplacer horizontalement pour déplacer le curseur de petite taille d'une barre de défilement.

L'*orientation* est une information exploitable pour les gestes. Par exemple, la sélection d'un élément dans les *Marking Menus* est basée sur cette propriété. Le geste permet également d'orienter un dispositif. Dans les *Motion Marking Menus* [Oakley et Park, 2009], l'inclinaison du dispositif de sa position horizontale à sa position verticale permet de sélectionner un élément du menu (Figure 4.8). Rap Menu [Ni et al., 2008] permet en orientant la paume de la main vers la gauche, en haut ou en bas, de sélectionner une sous-partie d'un menu circulaire. Dans la plupart des cas, l'orientation est quantifiable grâce aux informations générées par les diverses inclinaisons ou rotations. Cette caractéristique apparaît dans [Scoditti et al., 2011] où les auteurs décrivent les actions appliquées au dispositif avec, entre autres, les transformations affines impliquées dans le geste. Elle apparaît également dans [Ruiz et al., 2011] qui intègrent le nombre de dimensions impliquées dans le geste en différenciant les gestes avec un seul degré de rotation ou trois degrés de rotation.



FIGURE 4.8: Orientation du geste : le geste permet d'incliner le téléphone dans une direction pour sélectionner un élément du menu [Oakley et Park, 2009].

L'*accélération* et la *vitesse* du geste permettent de quantifier l'énergie transmise au dispositif. Baglioni et al. tirent parti de la signature accélérométrique des gestes pour différencier les gestes impulsifs des gestes fluides [Baglioni et al., 2009]. Les gestes impulsifs présentent un signal avec une forte variation d'amplitude pendant une courte durée de temps alors que les gestes fluides sont décrits avec des variations d'accélération de faible amplitude pendant une durée plus longue. Plus précisément, Ruiz et al. identifient trois niveaux

d'impulsion dépendant de la mesure de l'accélération [Ruiz *et al.*, 2011]. Ils définissent un niveau d'impulsion bas lorsque la mesure de l'amplitude de la secousse est inférieure à  $3\text{m.s}^{-3}$ , moyen pour une mesure comprise entre  $3\text{m.s}^{-3}$  et  $6\text{m.s}^{-3}$ , et haute pour une mesure supérieure à  $6\text{m.s}^{-3}$ .

La dimension vitesse est par exemple exploitée dans le geste de *Flick* pour insuffler la vitesse nécessaire pour faire défiler une liste [Baglioni *et al.*, 2011] ou envoyer un objet à un utilisateur à l'autre bout d'une table interactive [Wu et Balakrishnan, 2003]. Elle est aussi exploitée dans [Malacria *et al.*, 2010] où la vitesse de zoom est corrélée à la vitesse angulaire du mouvement circulaire effectué par le doigt.

#### 4.2.3.3 Combinaisons

Dans leur espace de conception, Wagner *et al.* [2013] distinguent des combinaisons de techniques d'interaction lorsque les tâches sont complexes. En nous inspirant de leurs travaux, nous considérons les combinaisons de gestes qui permettent de concevoir des techniques d'interaction plus expressives. Ces combinaisons peuvent être séquentielles ou concurrentes.



FIGURE 4.9: Combinaison de gestes : L'utilisateur effectue un geste de rotation du téléphone dans un sens. Puis il effectue un geste de rotation dans l'autre sens. Le geste a une signature unique qui permet d'être reconnu sans la nécessité d'utiliser un délimiteur [Ruiz et Li, 2011].

Les combinaisons séquentielles désignent deux ou plusieurs gestes qui sont réalisés les uns à la suite des autres. Par exemple, *DoubleFlip* [Ruiz et Li, 2011] est un geste composé d'un geste d'inclinaison d'axe longitudinal du dispositif mobile et son geste inverse pour repositionner le dispositif dans son état d'origine (Figure 4.9). Remarquons que ces gestes s'intègrent parfaitement et produisent un geste continu.

Cette continuité est une propriété intéressante qu'explore l'approche *CycloStar* [Malacria *et al.*, 2010]. Dans cette approche, les auteurs explorent les gestes de forme rectiligne et les gestes de forme elliptique. *CycloPan* est une technique d'interaction gestuelle qui permet de faire défiler une vue dans

toutes les directions de manière continue. Les gestes utilisés dans cette technique se composent d'une combinaison de gestes rectilignes d'aller-retour répétés autant de fois que nécessaire. De la même manière, *CycloZoom* repose sur un geste elliptique reproduit continuellement jusqu'à l'atteinte du niveau d'agrandissement attendu. Ces techniques reposent sur des gestes oscillatoires qui sont paramétrés par leur fréquence, leur amplitude, leur orientation ou leur position [Malacria *et al.*, 2010] (Figure 4.10).



FIGURE 4.10: Combinaison séquentielle : le geste elliptique continu permet de contrôler le zoom sur une carte affichée sur un tableau interactif [Malacria *et al.*, 2010].

Les combinaisons concurrentes de geste produisent des gestes complexes en réalisant deux ou plusieurs gestes simultanément. Ces gestes peuvent être produits de deux manières différentes. Tout d'abord, les gestes peuvent être produits simultanément par les deux mains. C'est le cas pour les gestes de pincement (*Pinch*) sur table interactive qui nécessitent d'écarter les deux mains en contact avec la surface pour contrôler le niveau d'agrandissement de l'objet manipulé. Dans [Bailly *et al.*, 2011], sélectionner un élément dans un menu exige de montrer les doigts des deux mains et fermer les deux mains simultanément pour valider.

Une autre manière est de combiner deux gestes avec la même main lors de la saisie d'un objet. *TiltText* requiert d'appuyer sur une touche du clavier téléphonique pour sélectionner un groupe de caractères et de réaliser un geste d'inclinaison du téléphone en même temps pour sélectionner un caractère [Wigdor et Balakrishnan, 2003]. Cette parallélisation de deux gestes peut

augmenter l'expressivité du vocabulaire gestuel tout en préservant les performances du geste le moins rapide.

#### 4.2.4 Contexte

Le *contexte* d'interaction peut influencer le choix du type de geste à réaliser. Ruiz *et al.* [2011] considèrent le contexte comme une situation qui entraîne un geste particulier comme le geste de rapprocher le téléphone de son oreille quand un appel arrive. Baglioni *et al.* [2009] suggèrent que l'environnement dans lequel évolue l'utilisateur peut perturber la réalisation des gestes. Par exemple, dans les transports en commun, l'espace disponible pour l'interaction est limité, ceci contraignant l'utilisateur à réaliser seulement des gestes de faible amplitude. Wobbrock *et al.* [2009] établissent une relation entre le geste et l'objet qui est manipulé. Ils différencient les gestes qui ont seulement besoin des informations de l'objet qui est manipulé en opposition aux gestes qui nécessitent des informations du monde dans lequel ils sont réalisés, ou des gestes indépendants de toute information. Scoditti *et al.* [2011] envisagent le contexte comme l'environnement des actions inconscientes qui nourrissent un modèle permettant d'agir au nom de l'utilisateur.

Dans notre approche, nous appelons contexte l'ensemble des facteurs susceptibles de faciliter la réalisation d'un geste. Ces facteurs peuvent être d'ordre *mécanique*, d'ordre *perceptif* ou *social*.

Comme le souligne Baglioni *et al.* [2009], en situation de mobilité, l'environnement peut devenir instable. Par exemple, les secousses liés au déplacement, à pied ou dans les transports en commun, peuvent dégrader la qualité des gestes effectués et ainsi affecter la reconnaissance.

L'attention visuelle peut être sollicitée si un retour visuel se produit pendant le geste spécialement si la tâche requiert une attention continue [Ruiz *et al.*, 2011]. Dans ce cas, les gestes mis à disposition pour effectuer la tâche doivent être conçus en tenant compte de ces contraintes. Par exemple, si un dispositif équipé d'un écran est soumis à des translations, la lecture des informations affichées peut s'avérer difficile.

Enfin, l'environnement social peut contribuer à l'acceptation ou le rejet des gestes (Figure 4.11). Par exemple, les gestes de petites amplitudes, métaphoriques ou familiers sont mieux acceptés socialement [Rico et Brewster, 2010]. Ces auteurs rapportent également que l'apparence perçue des gestes par les





FIGURE 4.11: Contexte social : étude de la préférence de réalisation des gestes dans un espace public [Rico et Brewster, 2010].

utilisateurs serait une caractéristique plus importante que la vitesse ou l'accélération transmise au dispositif par les gestes pour leur acception sociale.

#### 4.2.5 Conclusion

L'interaction gestuelle propose plusieurs classifications pour aider les concepteurs de techniques d'interactions. Cependant, la plupart de ces classifications sont spécifiques à un type d'interaction gestuelle particulier (interaction sur surface, interaction et interaction mains libres...) ou un contexte particulier (interaction avec un téléphone mobile, interaction avec une table interactive...). La classification proposée tente d'unifier les caractéristiques des gestes utilisés pour communiquer avec les systèmes dans la littérature en Interaction homme-machine. Inspirée par les caractérisations existantes, celle que nous proposons s'intéresse à la fonction du geste, son anatomie et le contexte dans lequel il est réalisé.

Nous remarquons que combiner différents types de gestes dans une même technique d'interaction amène plusieurs avantages tels qu'augmenter l'expressivité du geste en préservant des performances raisonnables. Dans le chapitre suivant, nous présentons une étude qui propose de tirer avantage de la combinaison de plusieurs types d'interaction pour augmenter l'expres-

sivité de la télécommande traditionnelle.

---

## TECHNIQUE D'INTERACTION GESTUELLE INSTRUMENTÉE POUR LA TÉLÉVISION

---

### 5.1 INTRODUCTION

Comme nous l'avons vu précédemment, les télécommandes traditionnelles connaissent de nombreuses limitations. Cependant, elles restent omniprésentes, peu onéreuses à fabriquer, et familières des téléspectateurs. Du fait de l'apparition de nouveaux services ou de nouveaux appareils, les télécommandes tendent à devenir de plus en plus complexes ou à voir leur nombre augmenter démesurément dans le salon. Réduire le nombre de boutons améliore l'utilisation des télécommande mais réduit d'autant leur expressivité qui se manifeste alors par des menus linéaires hiérarchiques toujours plus complexes. Comme nous l'avons vu précédemment dans le chapitre 3, une alternative est d'utiliser la modalité gestuelle pour augmenter la bande passante interactionnelle en étudiant l'orientation et la position de la télécommande.

Dans cette partie, nous nous intéressons à l'amélioration de la télécommande en exploitant les mouvements du dispositif et du poignet. Dans un premier temps, nous proposons une revue des travaux qui nous ont inspiré dans cette recherche, en particulier ceux de [Rahman \*et al.\* \[2009\]](#). Puis nous décrivons une étude exploratoire qui nous a aidé à comprendre quels sont les mouvements réalisés lorsque les utilisateurs manipulent une télécommande. Enfin, grâce aux résultats issus de cette étude, nous proposons de nouvelles techniques d'interaction pour faciliter l'accès aux éléments d'un système de menu.

### 5.2 TYPE DU MOUVEMENT DU DISPOSITIF

Dans le travail présenté ici, notre attention est centrée sur le mouvement du dispositif d'interaction dans l'espace. Pour une catégorisation de l'interaction



gestuelle, nous invitons les lecteurs à se référer au chapitre 4 de ce manuscrit.

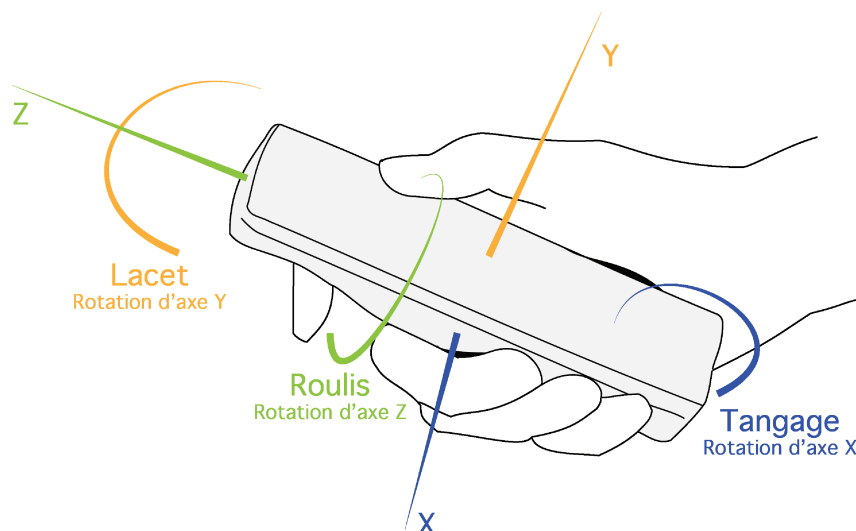


FIGURE 5.1: Illustration des degrés de liberté du mouvement du dispositif.

La mobilité d'un dispositif dans l'espace est exprimable par une combinaison de transformations spatiales selon six degrés de liberté (Figure 5.1). Ces six degrés incluent un mouvement translationnel sur trois axes différents (les axes X, Y et Z sur la figure 5.1) et un mouvement rotationnel sur chacun de ces trois axes. La rotation sur l'axe transversal X est appelée tangage, sur l'axe Y lacet, et sur l'axe longitudinal Z roulis.

L'utilisation de l'orientation du dispositif a largement été exploitée pour l'interaction à une main avec les dispositifs disposant d'un écran de taille limitée tels que les assistants personnels (PDAs) ou les téléphones mobiles. En 1993, [Fitzmaurice et al. \[1993\]](#) proposent de naviguer dans un espace de travail en trois dimensions à l'aide de *Chameleon*, un dispositif portable doté d'un écran de 4 pouces et de capteurs d'orientation. Ils présentent une nouvelle façon de naviguer dans une feuille de tableur cubique (*i.e.* en trois dimensions) (Figure 5.2). Les translations de gauche à droite du dispositif dans l'espace contrôlent le mouvement de la case du tableur, selon la colonne et la ligne, affichée au centre de l'écran. Rapprocher ou éloigner le dispositif de l'utilisateur active la navigation temporelle, offrant la possibilité de visualiser le contenu d'une même case à différents moments. Les auteurs suggèrent d'utiliser l'inclinaison du dispositif pour atteindre directement l'une des 8 cellules adjacentes à la case affichée au centre de l'écran du dispositif, devenant ainsi à son tour la cellule centrale. Ils évoquent également l'utilisation de l'in-

clinaison du dispositif pour exécuter des commandes. Enfin, ils envisagent d'utiliser l'intensité de l'inclinaison du dispositif pour contrôler la vitesse de déplacement durant la navigation.

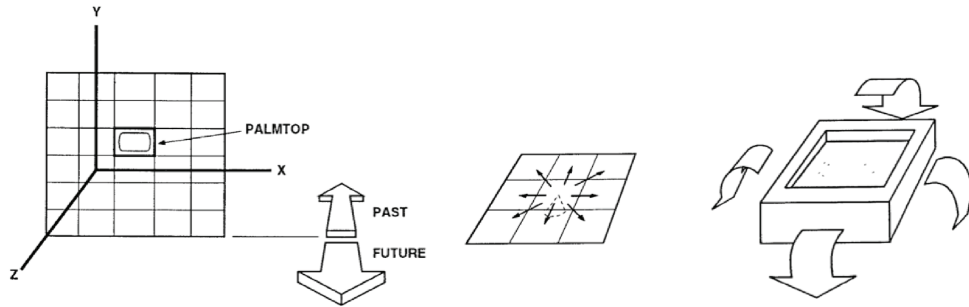


FIGURE 5.2: Chameleon : A gauche, les translations du dispositif permet de se déplacer dans la grille de cellule et le temps. A droite, l'inclinaison du dispositif permet de sauter de cellule adjacente en cellule adjacente.

Des techniques basées sur l'inclinaison d'un seul axe permettent d'interagir avec des appareils mobiles. *Rock 'n' Scroll* a introduit le changement automatique d'orientation de l'affichage sur l'écran du mode portrait au mode paysage [Bartlett, 2000]. Selon l'orientation de départ du dispositif, tenir le dispositif dans une nouvelle orientation pendant quelques secondes permet de passer d'un mode à l'autre. Notons que cette fonctionnalité est désormais intégrée dans la plupart des dispositifs mobiles doté d'un écran tactile. Dans *Double Flip*, une rotation rapide d'un smartphone dans un sens puis dans l'autre sens autour de son axe longitudinal (roulis) est utilisé comme délimiteur [Ruiz et Li, 2011]. Les *Motion Marking Menus* se basent sur l'amplitude du tantage pour sélectionner un élément dans un menu radial (Figure 5.3).

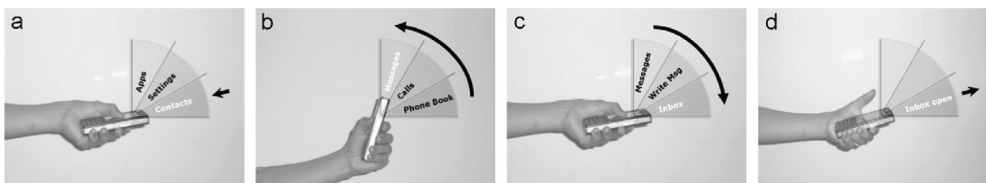


FIGURE 5.3: Marking Motion Menu. (a) L'utilisateur appuie contre l'écran tactile pour sélectionner le menu *contacts*. (b) Le menu *contacts* est affiché et l'utilisateur incline le dispositif pour sélectionner le sous-menu *messages*. (c) Le sous-menus *message* est affiché et l'utilisateur a repositionné le dispositif en position horizontale pour sélectionner la commande *inbox*. (d) Le contact avec l'écran est rompu pour actionner la commande.

Des techniques d'interaction plus complexes proposent de combiner deux types de rotation. Pour saisir du texte, *Unigesture* utilise le tangage et le roulis du dispositif pour sélectionner une des huit zones disponibles correspondant à une subdivision de l'alphabet en groupes de caractères [Sazawal *et al.*, 2002]. Lorsque le dispositif retourne dans la position initiale, la zone est sélectionnée. Le système infère ensuite le mot en analysant la séquence de rotations engagées pendant l'interaction. *Gestext* combine des rotations de roulis et de tangage pour sélectionner un caractère à l'aide de déplacement horizontaux et verticaux dans une matrice [Jones *et al.*, 2010]. Une séquence de rotations est nécessaire pour sélectionner un caractère en diagonale. *TiltText* propose de choisir un groupe de caractère en appuyant sur une touche du clavier numérique puis d'effectuer une rotation de type tangage ou roulis ou une combinaison des deux pour la saisie de texte sur téléphone mobile [Wigdor et Balakrishnan, 2003].

Les translations du dispositif ont également été explorés pour l'interaction avec les systèmes. *CubeKey* est une technique de saisie de texte pour écran distant où les lettres sont disposées dans une matrice cubique de longueur 3 (Figure 5.4) [Shoemaker *et al.*, 2010]. Le déplacement du dispositif mobile dans cet espace en trois dimensions permet de sélectionner une lettre contenue dans un sous-cube de la matrice. Dans la technique *Sweep*, les translations d'un téléphone portable contrôlent un curseur similairement à une souris optique [Ballagas *et al.*, 2005]. Le déplacement du curseur est calculé à partir du mouvement entre deux images successives capturées régulièrement par la caméra du téléphone.

Boring *et al.* [2009] comparent les mouvements de rotation et une combinaison de mouvements de translation et de rotation d'un téléphone mobile pour contrôler un pointeur sur de grands écrans à distance. Dans le premier cas, les mouvements de tangage sont associés au déplacement vertical et les mouvements de roulis au déplacement horizontal du curseur. Dans le second cas, inspiré par *Sweep*, les mouvements du curseur correspondent au mouvement de translation du téléphone. Les auteurs remarquent que les mouvements de translation horizontale peuvent être complétés par des mouvements de roulis quand l'amplitude du geste est importante. Lorsque c'est le cas, la vitesse du pointeur est augmentée. Les résultats de la comparaison sur une tâche de pointage montrent que les deux techniques sont propices aux erreurs de sélection à cause de dépassements de cible pour la technique basée sur les rotations, et de mouvements parasites lors de la sélection de la cible pour l'autre. Cela souligne l'importance de bien choisir le gain entre le mouvement du dispositif et le curseur, et d'autre part du rôle essentiel du délimiteur de geste

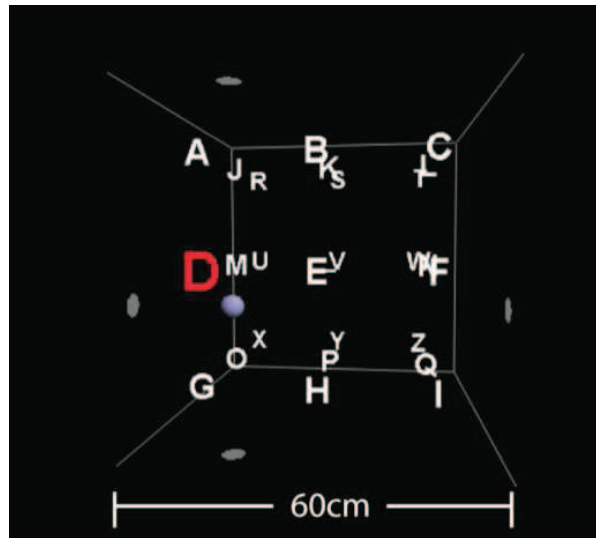


FIGURE 5.4: *CubeKey* : les mouvements du dispositif dans l'espace en 3 dimensions correspondent directement au mouvement du curseur à l'intérieur du cube.

comme expliqué au chapitre 4.

### 5.3 AMPLITUDE DU MOUVEMENT DU DISPOSITIF

La mesure de l'amplitude du mouvement peut servir au contrôle continu de paramètres. À la manière d'un joystick analogique, [Boring et al. \[2009\]](#) utilisent l'amplitude des rotations de tangage et roulis pour contrôler la vitesse du déplacement horizontal et vertical d'un curseur sur un écran distant. Plus l'amplitude de la rotation est importante et plus la vitesse de déplacement du curseur est rapide.

Le mesure de l'amplitude peut aussi permettre de sélectionner des modes. Par exemple, dans *TiltText*, les auteurs suggèrent d'utiliser l'amplitude des rotations pour distinguer la saisie des lettres en minuscules ou en majuscules [[Wigdor et Balakrishnan, 2003](#)]. Les auteurs de *Gestext* exploitent des rotations de "double profondeur" caractérisées par une amplitude importante. Les rotations de roulis simples permettent de sélectionner un caractère à gauche ou à droite du curseur de sélection alors que les rotations de double profondeur entraînent le mouvement du curseur de sélection vers le groupe de caractères à gauche ou à droite (Figure 5.5). Tenu dans la main, [Oakley et Park \[2009\]](#) proposent de diviser l'amplitude offerte par le tangage entre la position ho-

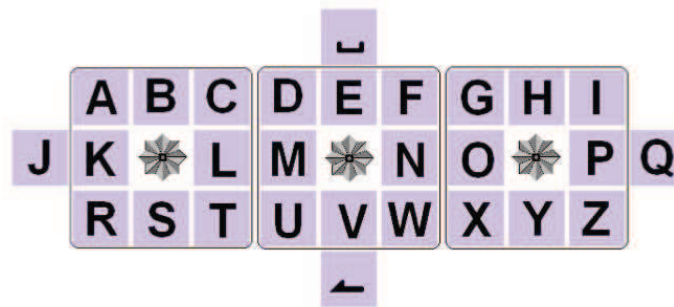


FIGURE 5.5: Gestext : disposition du clavier *tri-center*. Le curseur est initialement disposé au centre du clavier. Une simple rotation de roulis permet de sélectionner un caractère du bloc de gauche ou à droite. Une rotation de roulis de profondeur double sélectionne le centre du bloc de gauche ou de droite.

horizontale et verticale d'un dispositif. L'amplitude de la rotation permet de sélectionner jusqu'à 3 éléments différents par sens dans un menu sans retour visuel (Figure 5.3). Oakley et O'Modhain [2005] explorent également une situation dans laquelle l'amplitude de rotation est directement associée à un contrôle de position dans une liste d'éléments affiché sur l'écran d'un PDA. Les résultats suggèrent une limite acceptable de 15 éléments avec retour visuel pour cette technique dont l'amplitude s'élève à 90°.

Rahman *et al.* [2009] explorent un espace de conception des techniques d'interaction basées sur l'inclinaison du dispositif. En partant des capacités du mouvement du poignet, ils dressent un espace en s'appuyant sur les axes du mouvement, le type de contrôle (vitesse ou position), la discrétisation de l'amplitude de l'inclinaison, le mécanisme de sélection et le type de retour. Les résultats de leur diverses expérimentations montrent que, en tenant compte des contraintes du mouvement du poignet et avec une discrétisation quadratique, l'amplitude du tangage est confortablement divisible en 12 niveaux alors que l'amplitude de roulis est divisible en 16 niveaux d'environ 5° chacun. La discrétisation quadratique permet de diviser l'espace d'inclinaison de telle manière que les extrémités de l'amplitude du mouvement s'allouent une quantité importante de l'espace. Notons que ces résultats sont valides lorsqu'un retour visuel est offert aux utilisateurs.

Cette revue de la littérature permet de tirer plusieurs constats. Tout d'abord, l'utilisation des translations et des rotations est fortement inspirée par des métaphores d'interaction. Par exemple, dans *Sweep*, la technique de déplacement de curseur à distance sur une surface verticale par translation du dispositif

est comparée au comportement d'une souris optique [Ballagas *et al.*, 2005]. Dans [Boring *et al.*, 2009], la vitesse de déplacement du curseur dépend de l'amplitude, ce qui rappelle le comportement d'un joystick. Un autre constat est que le lacet a rarement été exploité jusqu'à présent. Comme nous l'avons vu à la section 3.2.1.1, une première raison plausible est que cette rotation n'est pas facilement mesurable avec les accéléromètres sur lesquels se basent la plupart des travaux antérieurs. Une autre raison est la désorientation visuelle peu commode pour l'interaction qu'entraîne le lacet sur les dispositifs dotés d'un écran. Dans le cas où l'affichage est déporté, le lacet reste donc exploitable. Enfin, les travaux précédents offrent des recommandations dans le cas où le retour visuel est disponible [Rahman *et al.*, 2009]. Le cas de l'interaction "eyes free" qui est important dans notre contexte d'usage car on souhaite éviter à l'utilisateur d'avoir à regarder sa télécommande, ne semble pas avoir été considéré dans la littérature.

Nous proposons d'étudier comment les mouvements du dispositif peuvent améliorer l'interaction en utilisant une télécommande dans le cadre de la télévision interactive. Dans une première étude exploratoire, nous observons que la localisation de l'affichage (embarqué sur le dispositif ou sur un écran déporté) a une influence sur les rotations utilisées. Dans un second temps, nous explorons comment exploiter la rotation non utilisée par la première tâche. Les résultats de ces études alimenteront un espace de conception de technique d'interaction visant à fournir un accès rapide et facile aux éléments d'un menu pour la télévision. Pour finir, nous présentons une évaluation comparative de ces techniques.

#### 5.4 CORRESPONDANCE ENTRE LES GESTES ET LE MOUVEMENT DU DISPOSITIF

La mobilité de la télécommande dans un espace 3D est une combinaison de transformations spatiales selon six degrés de liberté (Figure 5.1). Néanmoins, l'utilisation de tous ces degrés et leurs combinaisons n'a pas été encore exploré. Par ailleurs, nous nous questionnons sur l'effet du contexte d'utilisation sur ces choix. Dans cette section, nous rapportons une première expérimentation qui répond à ces interrogations.

### 5.4.1 Expérimentation

Comme nous l'avons remarqué précédemment, de nombreuses techniques d'interaction sont basées sur les rotations de type tangage et roulis lorsque l'écran est embarqué dans le dispositif mais aussi lorsqu'il est déporté [Partridge *et al.*, 2002; Sazawal *et al.*, 2002; Wigdor et Balakrishnan, 2003; Ballagas *et al.*, 2005]. Par exemple, dans *GestText* une sélection à gauche ou à droite sur un écran distant s'effectue par le biais d'une rotation de roulis du dispositif [Jones *et al.*, 2010].

Cette observation nous conduit à notre première expérimentation inspirée par les travaux de Wobbrock *et al.* [2009] sur les gestes définis par les utilisateurs. Ces travaux proposent de concevoir un vocabulaire gestuel pour des réaliser les tâches courantes sur table interactive en demandant aux participants quel geste pourrait être la cause des effets d'un ensemble d'actions qui leurs sont présentées. Appliqués à notre contexte, et en considérant la localisation de l'affichage comme facteur, notre objectif est de déterminer quelles sont les rotations utilisées par les participants pour effectuer une sélection directionnelle. Notre hypothèse est que, spontanément, les utilisateurs ne font pas appel aux mêmes gestes selon que l'affichage est déporté devant eux ou intégré au dispositif qu'ils manipulent (Figure 5.6).

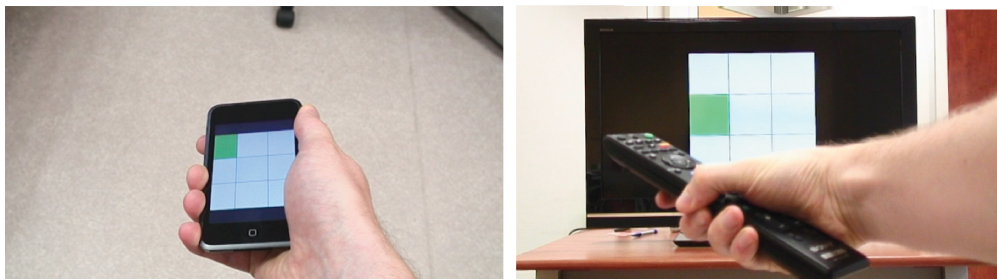


FIGURE 5.6: Expérience 1. A gauche, l'utilisateur incline le dispositif mobile pour atteindre la cible. A droite, l'utilisateur incline la télécommande pour atteindre la cible.

#### 5.4.1.1 Tâche

Cette expérimentation propose une tâche de sélection directionnelle (selon les directions cardinales et diagonales) réalisée à l'aide de gestes en l'air soit avec une télécommande et un écran déporté, soit avec un dispositif mobile équipé d'un écran. L'élément cible apparaît dans une grille de 3 cases par 3 où la cellule centrale reste toujours vide. Cette configuration permet d'identi-



fier chaque élément par sa position dans la grille. Les participants sont libres de réaliser des translations ou des rotations du dispositif pour atteindre la cible. Pour sélectionner un élément, il faut donc réaliser une translation vers la position de l'élément ou bien une rotation dans la direction de l'élément.

#### 5.4.1.2 *Protocole expérimental*

Pour chaque essai, une cellule de la matrice a été illuminée aléatoirement. Similairement à l'approche proposée dans [Wobbrock *et al.*, 2009], nous avons demandé aux participants d'effectuer le geste 3D qui avait, d'après eux, entraîné l'illumination de la cellule concernée. Par ailleurs, nous leur avons aussi expliqué que la cellule centrale correspond à la position de repos et de départ du geste. Nous leur avons finalement demandé de revenir à la position de repos après leur geste avant de commencer l'essai suivant. A des fins d'analyse, les participants ont été filmés durant l'expérimentation.

Nous avons considéré un facteur avec deux possibilités. La première condition consistait à interagir à l'aide d'un dispositif équipé d'un écran comme un téléphone portable. La seconde condition impliquait d'interagir avec un dispositif et un écran déporté, à l'image d'un téléspectateur interagissant avec sa télévision grâce à une télécommande. Dans la première condition, le stimulus est apparu sur l'écran intégré au dispositif, dans notre cas un iPod Touch. Dans la seconde condition, le stimulus est apparu sur l'écran déporté, une télévision haute définition, située à 3 mètres des participants. Afin de se rapprocher d'une situation en condition réelle, nous avons demandé aux participants de ne pas regarder la télécommande dans cette condition.

18 participants recrutés dans notre institution et âgés de 24 à 30 ans ont été divisés en groupe de 9 participants pour chaque condition. Nous avons recolté 8 gestes par participant correspondants à la sélection de chaque cellule de la matrice.

Les gestes ont été identifiés à partir des enregistrements vidéos parmi l'une des 6 catégories suivantes : translation sur l'axe X, l'axe Y, l'axe Z, tangage, roulis ou lacet (Figure 5.1). Pour nous assurer de la fiabilité des résultats, deux expérimentateurs ont procédé à la classification des gestes et confronté leurs résultats, similaires dans les deux cas. L'absence de mesures numériques permet seulement de reporter des tendances générales du fait de l'impossibilité de détecter des petits mouvements qui seraient indiscernables à l'œil nu.

### 5.4.1.3 Résultats

Sur les 9 participants qui ont utilisé l'iPod, 7 participants ont utilisé des rotations de tangage et de roulis. En ce qui concerne l'autre condition, nous avons observé 7 participants utiliser des rotations de tangage et de lacet. Le test exact de Fisher montre que ce facteur a un effet significatif sur ces résultats. Enfin 2 utilisateurs ont utilisé des mouvements de translation dans les deux conditions.

Aucun des participants n'a réalisé une combinaison de translation et de rotation. Les gestes réalisés sont tous des combinaisons de gestes impliquant soit des translations, soit des rotations. Leur comportement est resté consistant tout au long de l'expérimentation. 77,78% des participants ont spontanément effectué des rotations et seulement 22,22% des translations. Ce résultat, explicable par le fait qu'il est plus facile d'effectuer des rotations, les bras étant composés de plusieurs articulations [Hogan, 1985], nous a amené à étudier plus en profondeur les rotations dans la suite de ce travail.

Confirmant notre hypothèse, les participants n'ont pas effectué les mêmes rotations dans les deux conditions. Plus précisément, ils ont utilisé le tangage dans les deux mais le roulis dans la condition smartphone et les gestes de lacet dans la condition télécommande. Dans la condition smartphone, les participants ont utilisé une combinaison de roulis et de tangage permettant ainsi de garder l'écran du dispositif visible tout au long de l'interaction. Concernant le cas de la télécommande, les participants ont utilisé le dispositif comme si ils pointaient la cible présentée à l'écran en utilisant une combinaison de tangage et lacet. Ceci est en contradiction avec les travaux de Jones *et al.* [2010] qui avaient utilisé le roulis pour effectuer des déplacements horizontaux du curseur pour sélectionner les lettres d'un clavier affiché sur une télévision. En pratique, cela signifie qu'il est souhaitable que les télécommandes augmentées ne disposent pas seulement d'un accéléromètre mais également d'un gyroscope pour détecter efficacement les rotations de lacet comme le propose déjà Movea<sup>1</sup> avec la télécommande de la Freebox Révolution<sup>2</sup>.

Enfin, aucune translation sur l'axe des Z ou de rotation de type roulis n'a été constatée dans la condition avec télécommande. Cette observation indique que ces deux dimensions restent disponibles pour adresser des tâches supplémentaires. Néanmoins, la translation sur l'axe des Z entraîne la coordination motrice des articulations de l'épaule et du coude. De ce fait, cette coordination peut-être inconfortable lorsque l'utilisateur est assis dans son canapé.

---

1. <http://www.movea.com/applications/interactive-tv>

2. <http://www.free.fr/adsl/freebox-revolution.html>

Pour cette raison, dans la suite de ce travail, nous nous sommes intéressés uniquement à l'exploration du roulis.

## 5.5 DE L'UTILISATION DU ROULIS

Le roulis (Figure 5.7), résultant du mouvement de supination (rotation du poignet entraînant une posture de la main avec la paume vers le haut) et pronation (rotation du poignet entraînant une posture de la main avec la paume vers le bas) du poignet, a été précédemment étudié dans quelques travaux déjà évoqué à la section 5.3. Alors que [Rahman et al. \[2009\]](#) ont montré que les utilisateurs étaient capables de sélectionner 16 éléments à l'aide du roulis en présence de retour visuel, d'autres travaux se limitent à 3 éléments comme dans [\[Oakley et Park, 2007\]](#) qui présentent une technique avec un menu non visuel. Cette différence s'explique par l'absence d'affichage dans le second cas qui limite considérablement la précision. Cependant, nous pensons que les utilisateurs devraient être tout de même capables de sélectionner un nombre plus conséquent de commandes avec une précision raisonnable dans ce cas. Afin de déterminer cette limite, nous avons conduit deux nouvelles expérimentations pour explorer la résolution du geste de roulis en l'absence de retour visuel. Cette situation correspond au cas familier où l'utilisateur regarde la télévision et souhaiterait autant que possible éviter d'avoir à regarder la télécommande pour interagir.

### 5.5.1 Limite maximale de sélection et posture

Dans cette étude, nous nous intéressons à deux facteurs : le nombre d'éléments sélectionnables grâce au roulis et l'influence de la posture sur cette limite. Nous appelons posture les différentes configurations qui pourraient contraindre le mouvement du dispositif. Les mouvements de roulis impliquent plusieurs articulations, comme le poignet, le coude ou encore l'épaule, notamment lorsque les gestes sont de grande amplitude. Plus particulièrement, nous sommes intéressés par les situations où l'amplitude de mouvement des articulations du coude et de l'épaule est contrainte. Un exemple d'une telle situation est lorsque le téléspectateur est accoudé sur son canapé. L'avant bras posé sur l'accoudoir empêche le mouvement de l'articulation du coude et de l'épaule et diminue ainsi l'amplitude totale de mouvement disponible pour la télécommande.

Notre hypothèse est que les participants arriveront à discriminer au moins 7 éléments et jusqu'à 9 éléments en l'absence d'attention visuelle. Afin de vérifier cette hypothèse, 4 ensembles de tailles différentes avec 5, 7, 9 et 11 éléments ont été considérés. Notons que le cas avec 3 éléments a déjà été abordé dans [Oakley et Park, 2007]. Nous supputons aussi que la posture peut avoir une influence sur le nombre d'éléments que les utilisateurs sont capables de sélectionner. Lorsque l'utilisateur est dans son canapé avec une télécommande traditionnelle, la liberté de mouvement de ses articulations peut être plus ou moins importante selon la posture dans laquelle il se trouve. Afin de mesurer cet effet, nous avons considéré plusieurs postures durant cette étude :

- Posture #1 : seulement le poignet est libre. L'avant-bras est en repos sur la cuisse.
- Posture #2 : l'avant-bras est en contact forcé avec le long du corps, ceci permettant seulement la liberté du coude et du poignet.
- Posture #3 : aucune contrainte, les trois articulations sont libres de tout mouvement.

#### 5.5.1.1 *Discrétisation du mouvement de roulis*

Quantifier le mouvement de roulis revient à mesurer l'amplitude entre la position angulaire initiale et une position angulaire finale suivant l'axe longitudinal du dispositif. Dans notre expérimentation, nous avons choisi d'utiliser une *Wiimote*. Un avantage de la *Wiimote* est qu'elle est équipée de boutons judicieusement placés, comme par exemple une gâchette positionnée sous l'index ou les touches directionnelles facilement accessibles par le pouce quand l'utilisateur tient la télécommande. D'ailleurs, la gâchette nous servira de délimiteur gestuel pour cette expérimentation.

Pour mesurer l'amplitude angulaire, nous avons suivi la stratégie *Key Tilt* décrite dans [Wigdor et Balakrishnan, 2003] ou également appelé *Click-Tilt-Release* dans [Rahman et al., 2009]. Dans cette approche, l'utilisateur doit d'abord appuyer sur la gâchette avant de réaliser le mouvement de roulis. A la pression de la gâchette, le système enregistre la valeur initiale de l'orientation de la télécommande. Tant que la gâchette reste enclenchée, l'utilisateur effectue le mouvement de roulis. Au relâchement de la gâchette, le système enregistre la valeur angulaire résultant de l'orientation de la télécommande à la fin du geste. L'amplitude mesurée est la différence entre ces deux valeurs angulaires. La procédure est illustrée par la partie droite de la (Figure 5.7).

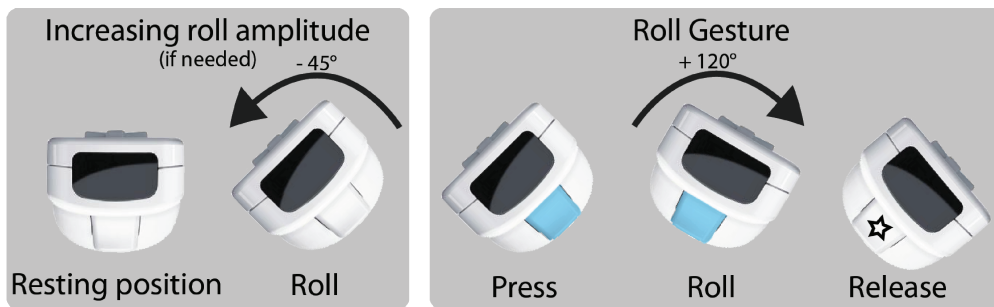


FIGURE 5.7: Exemple d'un geste de roulis. A gauche, un geste préparatoire pour augmenter l'amplitude disponible pour effectuer le geste. A droite, l'exécution du geste de roulis. La pression de la gâchette délimite le geste.

Cette stratégie avantageuse permet d'étendre la plage angulaire disponible pour la réalisation du geste. Le mouvement de roulis du dispositif est réalisé grâce à un mouvement de supination ou de pronation du poignet qui offre une portée totale de  $125^\circ$  [Rahman *et al.*, 2009]. Cette valeur peut être augmentée grâce au mouvement du coude et au mouvement de l'épaule. Par exemple, pour réaliser un geste de roulis avec une amplitude conséquente dans le sens horaire, le geste effectif peut être amorcé par un geste de roulis dans le sens anti-horaire avant d'appuyer sur la gâchette. Ceci permet d'augmenter la plage angulaire totale disponible pour le geste effectif. Ainsi, l'amplitude maximale de la rotation angulaire peut atteindre jusqu'à  $200^\circ$  et même plus dans certains cas, selon l'utilisateur et la posture (Figure 5.7).

#### 5.5.1.2 Tâche

Pour vérifier nos hypothèses nous avons demandé aux participants de réaliser des gestes de roulis de la télécommande. Les consignes données aux participants étaient de réaliser ces gestes le plus précisément et le plus rapidement possible sans regarder la télécommande en utilisant la stratégie *Click-Tilt-Release*. L'amplitude du geste à réaliser a été laissée à l'appréciation des participants. Nous avons par contre varié le nombre d'éléments qui constituait un des facteurs de l'expérience, en considérant les valeurs suivantes : 5, 7, 9, 11 éléments.

#### 5.5.1.3 Stimulus et retour

Le stimulus présenté aux participants se composait d'un ensemble de carrés de même taille alignés horizontalement (Figure 5.8). Le nombre de carrés affichées à l'écran correspondait au nombre d'élément de chaque ensemble

de notre facteur. Pour chaque essai, un élément parmi les éléments présentés à l'écran était coloré. Dans le cas où l'élément central était coloré, aucune rotation n'était nécessaire, le participant n'ayant seulement besoin que d'appuyer et de relâcher la gâchette. La position de chaque élément apportait deux informations : la direction et l'intensité du mouvement de roulis à réaliser.

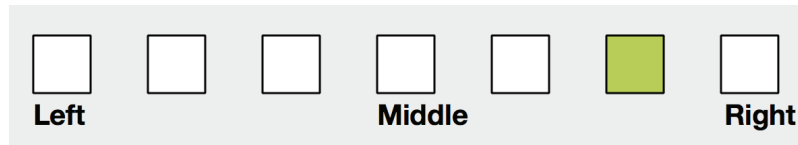


FIGURE 5.8: Stimulus pour la sélection d'une cible parmi sept en utilisant un geste de roulis.

Afin d'éviter de conditionner les participants, nous n'avons donné aucune information numérique ou symbolique sur la distance angulaire à parcourir. Cette neutralité de la représentation impose aux participants d'interagir sans conditionnement lié au stimulus. En pratique, plus l'élément coloré est éloigné de l'élément central et plus l'amplitude du geste de rotation réalisé est importante. Nous avons utilisé un écran de télévision pour afficher les stimulus.

Aucun retour visuel, audio ou haptique sur le mouvement n'a été fourni aux participants. Ceux-ci ont été obligés de se fier à la proprioception, la perception du mouvement du poignet et de l'avant bras, pour évaluer l'amplitude nécessaire pour effectuer le geste. Notre but était de mesurer le nombre maximum d'éléments discriminables et l'amplitude angulaire dont les participants avaient besoin pour discriminer chaque élément.

#### 5.5.1.4 *Participants et apparatus*

Neuf volontaires âgés de 22 à 28 ans dont 8 de sexe masculin ont été recrutés dans notre institut. L'expérimentation a été conduite en utilisant un *Macbook Pro* et un écran de télévision haute définition de 40 pouces. Les participants étaient assis dans un canapé disposé à environ 3 mètres de la télévision pour simuler un environnement domestique. Comme précisé précédemment, la télécommande utilisée était une *Wiimote*. Afin de communiquer avec la *Wiimote*, nous avons implémenté une adaptation de la librairie *Darwin Remote*<sup>3</sup>. Les applications qui ont servi dans cette expérimentation ont été développées

3. <http://darwiin-remote.sourceforge.net>

en utilisant la librairie Qt<sup>4</sup>.

#### 5.5.1.5 Procédure

Pour chaque participant, la session a débuté par un entraînement d'une série de mouvements avec 3 éléments. Chaque participant a complété 5 séries pour chaque ensemble d'éléments et de posture. Au sein d'une série, chaque élément est apparu une seule fois dans un ordre aléatoire. Les 4 ensembles de 5, 7, 9 et 11 éléments ont été présentés dans un ordre croissant, augmentant l'indice de difficulté à chaque série. Pour inhiber l'effet lié à la posture, les différentes configurations ont été ordonnées à l'aide d'un carré latin. Au total, nous avons récolté 4320 gestes répartis sur 9 participants, 4 ensembles d'éléments, 5 séries et 5, 7, 9 ou 11 essais en fonction de la taille de l'ensemble. Entre chaque essai, une pause de 1,5 seconde était imposée pour que le participant puisse replacer son poignet à la position de repos avant le prochain essai.

#### 5.5.1.6 Résultats

Les participants n'ayant (volontairement) pas reçu d'indication à propos de l'amplitude de mouvement à réaliser, d'importantes différences ont été observées. Par exemple, la plage angulaire entre l'amplitude minimale et maximale parmi tous les participants variait de 162° à 326° dans la condition de l'ensemble à 7 éléments. Notons que l'amplitude pouvait atteindre jusqu'à 326° grâce à la stratégie *Click-Tilt-Release*.

Sans surprise, la différence entre l'amplitude minimale et l'amplitude maximale dépend très significativement de la taille de l'ensemble d'éléments ( $F_{3,24} = 97,47, p < 0,0001$ ) et augmente avec le nombre d'éléments. La différence moyenne est de 159° pour 5 éléments, 306° pour 7 éléments, 343° pour 9 éléments et 371° pour 11 éléments toutes postures confondues. La dépendance de l'amplitude maximale avec la posture est faible mais significative ( $F_{3,24} = 9,8, p < 0,01$ ), avec en moyenne 273° pour la posture #1, 290° pour la posture #2, et 321° pour la posture #3.

Le tableau 5.1 montre les variations angulaires moyennes pour chaque élément selon la taille du groupe d'éléments. Les participants ont ajusté l'amplitude de leur mouvement de roulis proportionnellement au nombre d'éléments contenu dans l'ensemble. Une corrélation linéaire très importante et

---

4. <http://qt-project.org>



		Identifiant de cible										
		-5	-4	-3	-2	-1	0	1	2	3	4	5
Nombre d'éléments	5 elts				-81,2	-43,6	0,3	42,8	78			
	7 elts			-155	-90,3	-49,6	-0,9	52,9	88,6	150,2		
	9 elts		-173	-116	-83,1	-51,3	-0,6	50,9	81,9	118,5	171	
	11 elts	-188	-136	-101	-77,5	-46,8	-0,6	46,4	70,6	99,3	138,8	184,4

TABLE 5.1: Variation angulaire moyenne pour chaque élément par taille de groupe

une ordonnée à l'origine proche de  $0^\circ$  pour chaque groupe est remarquable ( $r^2 > 0,995$  pour tous les ensembles et toutes les ordonnées à l'origine étant inférieure à  $1^\circ$  et non significative).

Pour évaluer la reconnaissance des différentes amplitudes de rotation réalisées pour sélectionner un élément, nous avons mis en œuvre un algorithme du plus proche voisin avec les 5 voisins les plus proches [Cover et Hart, 1967]. Pour mesurer la distance entre les voisins, nous calculons la distance euclidienne entre la valeur de l'amplitude angulaire de l'échantillon à évaluer et la valeur de l'amplitude angulaire de l'échantillon dans la base. Cet algorithme a pour avantage d'être simple, tolérant au bruit sur la classification et d'être efficace pour un temps de calcul raisonnable [Bailly *et al.*, 2008]. La base d'évaluation et la base d'apprentissage ont été séparées en utilisant la méthode de validation croisée où chaque élément est comparé à tous les éléments de la base d'apprentissage privée de cet élément. Les résultats montrent que la précision diminue significativement lorsque la taille du groupe augmente ( $F_{3,24} = 53,52, p < 0,0001$ ). Pour une taille de 5 éléments, la précision est de 96,3%, pour 7 éléments 87,7%, pour 9 éléments 77,1%, pour 11 éléments 63,1%. Aucun effet n'a été observé pour le facteur de posture.

#### 5.5.1.7 Discussion

La présence d'un retour visuel continu sur le mouvement angulaire influence la relation entre la variation angulaire et le groupe d'éléments. Alors que Rahman *et al.* [2009] ont observé une relation quadratique, nos résultats révèlent une relation linéaire lorsque les gestes sont réalisés sans retour visuel.

Étonnamment, la posture n'a pas d'influence sur la reconnaissance des gestes alors que son effet sur l'amplitude maximale de variation est signi-

ficatif. Les participants divisent l'amplitude disponible en des secteurs quasi égaux dont la taille dépend de la posture. Ce résultat est intéressant car il suggère un patron invariant facilement reconnaissable.

Les participants ont été capables de discrétiser 5 éléments avec des mouvements de roulis avec des taux de reconnaissances supérieurs à 95%. La difficulté des conditions avec l'ensemble de 9 et 11 éléments sans retour visuel était trop élevée pour les participants. Enfin, les résultats issus de la condition avec l'ensemble à 7 éléments donne un taux de reconnaissance un peu inférieur à 90%. Cette valeur un peu limite nous a conduit à mener une expérimentation supplémentaire pour cette condition.

#### 5.5.2 *Approfondissement du cas à 7 éléments*

Pour cette nouvelle expérimentation, la valeur des amplitudes du mouvement à réaliser a été affichée aux participants. Ces amplitudes étaient  $-135^\circ$ ,  $-90^\circ$ ,  $-45^\circ$ ,  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ . Ces valeurs ont été sélectionnées suite aux résultats des expérimentations précédentes, l'incrément étant juste arrondi jusqu'à  $45^\circ$  pour faciliter la représentation mentale angulaire auprès des participants. Dans cette expérimentation, l'effet de la posture n'a pas été étudié.

Une vingtaine de nouveaux participants âgés de 25 à 31 ans et droitiers ont pris part à l'expérimentation. Pendant 5 minutes, une première phase d'apprentissage a permis aux participants de se familiariser avec le geste de roulis grâce à l'affichage textuel continu de l'angle couvert par le mouvement. Par la suite, lors de la phase de recueil des données, les participants ont été privés de cette information visuelle. Chaque participant a réalisé quinze séries où chaque élément parmi les sept disponibles sont apparus une seule fois dans un ordre aléatoire. Pour résumer, le plan expérimental a impliqué 20 participants avec 15 séries de 7 éléments pour totaliser 2100 sélections.

##### 5.5.2.1 *Résultats*

Comme précédemment, la reconnaissance des éléments sélectionnés a été évaluée avec la méthode des  $k$  plus proches voisins, avec 5 voisins et la technique de validation croisée. Le taux de reconnaissance obtenu est de 91%, ce qui constitue une valeur acceptable pour une utilisation réaliste. Nous nous sommes intéressés dans un second temps à l'analyse des données pour chaque participant. Cette situation correspond à une personnalisation du système dans laquelle la base d'apprentissage du reconnaisseur est alimentée

Identifiant de cible	-3	-2	-1	0	1	2	3
Direction	Gauche				Droite		
Temps de sélection moyen (en sec)	3,53	2,67	2,44	1,40	2,60	2,98	3,27
Amplitude angulaire moyenne (en °)	-134	-85	-51	0	49	86	127

TABLE 5.2: Temps moyen de sélection (en secondes) et amplitude angulaire moyenne pour chaque élément (en degrés).

avec les données d'un utilisateur spécifique. Le taux moyen de reconnaissance obtenu dépasse 96%.

Sans surprise, le temps de sélection a été significativement dépendant de la position de l'élément à sélectionner ( $F_{6,114} = 32,4, p < 0,001$ ) (Tableau 5.2). En moyenne, le temps de sélection pour l'élément central (1,40 secondes) a été significativement plus rapide que pour l'élément -1 (2,44s) et +1 (2,60s), eux-même significativement plus rapide que pour les éléments -3 (3,53s) et +3 (3,27s). Enfin, le temps de sélection pour l'élément central a été significativement plus rapide que pour les éléments -2 (2,67s) et +2 (2,98s).

Le tableau 5.2 présente la valeur des angles qui nous serviront à paramétrer les techniques basées sur les mouvements de roulis dans l'expérience suivante. Ces valeurs sont les valeurs moyennes des amplitudes angulaires des participants pour chaque élément.

#### 5.5.2.2 Synthèse

Notre investigation sur la sélection d'éléments à l'aide de gestes de roulis a révélé que l'approche est plutôt robuste, la posture n'ayant aucun effet sur le taux de reconnaissance. Il s'avère que tous les participants ont pu facilement sélectionner 5 éléments et voire même plus en particulier lorsque l'apprentissage est personnalisé. Au vu de ces résultats, nous pouvons conclure que 7 éléments constitue la limite maximale de discrétisation de l'espace angulaire fourni par le geste de roulis lorsque les utilisateurs n'ont pas de retour visuel et ne regardent pas la télécommande.

### 5.6 CONCEPTION DE NOUVELLES TECHNIQUES D'INTERACTION

A partir des résultats obtenus précédemment, nous proposons maintenant de nouvelles techniques d'interaction pour interagir à l'aide d'une télécom-

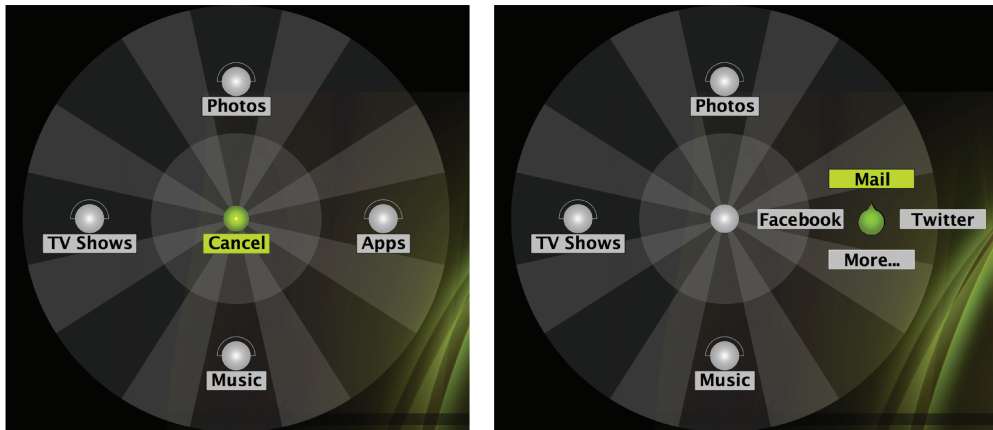


FIGURE 5.9: Un exemple de mode novice de *Marking Menu* à deux niveaux avec 4 éléments affiché sur l'écran de la télévision.

mande. Les techniques proposées sont inspirées des *Marking Menus* [Kurtenbach et Buxton, 1991]. Plus précisément, ces techniques permettent d'interagir avec un menu hiérarchique à deux niveaux. Chacune de ces techniques a pour objectif de sélectionner un élément dans chacun des niveaux du menu et fonctionne dans un mode novice où le menu est affiché sur l'écran de la télévision et un mode expert sans affichage du menu (Figure 5.9). Dans un premier temps, nous allons nous intéresser au mode expert. Puis nous aborderons la conception du mode novice.

Même si l'affichage du menu sera de toute façon nécessaire pour exécuter des commandes peu fréquentes, il est avantageux de permettre aux utilisateurs de réaliser les actions les plus courantes ou d'accéder aux contenus favoris aisément et rapidement tout en leur épargnant la navigation dans un menu complexe. Les résultats précédents ont montré que les rotations de tangage, roulis et lacet sont exploitables dans notre contexte. Les télécommandes sont dotées de boutons, ce qui ajoute encore une dimension interactive supplémentaire.

Les gestes de tangage et lacet correspondent à deux rotations différentes. Ces deux dimensions sont intégrables au sens de Jacob *et al.* [1994] car la structure perceptuelle de la tâche s'accorde avec la structure de la combinaison de ces rotations. Elles peuvent être contrôlées simultanément et un geste composite fait sens pour l'utilisateur (i.e déplacer le curseur en diagonale). Ici, la combinaison de ces rotations permet de sélectionner une direction comme nous l'avons vu dans la première expérimentation de ce chapitre. Choisir une direction pour sélectionner une commande rappelle le mécanisme des

*Marking Menus* dont la représentation graphique reflète l'espace de contrôle du dispositif [Jacob *et al.*, 1994]. Il est intéressant de remarquer que *GesText* combine des rotations de tangage et de roulis pour déplacer le curseur de sélection horizontalement et verticalement [Jones *et al.*, 2010]. Dans cette technique, les degrés ne sont pas intégrables mais séparables : le mouvement du curseur nécessite deux types de rotations différents qui ne sont pas effectués en un seul geste.

Les *Marking Menus* ont prouvé leur efficacité pour des gestes de translation à deux dimensions sur des surfaces planes et rigides, notamment pour l'interaction sans retour visuel. Ces résultats nous laissent supposer que transposer cette technique avec des gestes de rotation dans un espace 3D pourrait être prometteur. Remarquons enfin que cette transposition se rapproche davantage des *Marking Menus* originaux que des techniques comme les *Marking Motion Menus* [Oakley et Park, 2007]. En effet, ces derniers sont basés sur la pronation et supination du poignet pour contrôler le tangage entre une position horizontale et verticale du dispositif.

Nous considérons également les gestes de roulis comme expliqué dans les parties précédentes.

Les télécommandes traditionnelles sont en général pourvues d'un pavé directionnel comprenant quatre boutons. Ce dispositif est adapté pour les tâches directionnelles lorsqu'elles ne sont pas trop répétitives comme dans le cas d'une navigation dans un menu complexe. L'interaction avec ces boutons est la troisième dimension que nous avons retenue. La nature directionnelle commune de cette dimension avec la combinaison des gestes de tangage et lacet est intéressante car elle permet une comparaison directe. En outre, il semble également possible de combiner l'interaction à l'aide de boutons avec les gestes en 3 dimensions. La combinaison de ces dimensions pourrait être performante en temps comparée à des techniques n'utilisant que deux dimensions et nécessitant donc d'effectuer la sélection en deux étapes.

Chaque niveau du menu pouvant être contrôlé par une des trois dimensions, ceci donne une possibilité de neuf combinaisons différentes (tableau 5.3). Dans la suite, nous listons et évaluons l'intérêt de ces combinaisons.

*Tangage-Lacet (T&L) / T&L*. Cette combinaison correspond à un portage des menus *Multi-Strokes* [Zhao et Balakrishnan, 2004]. Les utilisateurs doivent compléter deux gestes directionnels successifs. Chaque geste directionnel est

		Second geste			
Premier geste		Tangage et Lacet		Roulis	Bouton directionnel
	Tangage et Lacet	Menu multi-stroke avec gestes 3D		Parallélisation des gestes 3D	Nombre d'actions minimales important
	Roulis	Contrainte mécanique	bio-importante	Temps d'exécution trop important	Nombre d'actions minimales important
	Bouton directionnel	Parallélisation des gestes 2D et 3D		Parallélisation des gestes 2D et 3D	Technique de référence

TABLE 5.3: Considération des techniques de l'espace de conception pour évaluation : les techniques en fond gris ne sont pas retenues pour évaluation.

délimité par l'appui et le relâchement de la gâchette.

*T&L / Roulis.* La combinaison du geste de T&L et du geste de roulis entraîne un mouvement fluide qui peut être considéré comme un geste continu. De ce fait, il nous a paru intéressant de considérer cette combinaison.

*Roulis / T&L.* La combinaison inverse a produit des résultats décevants par rapport aux autres combinaisons lors des évaluations préliminaires. Les gestes de roulis diminuent la portée disponible et rend inconfortable les gestes T&L. Pour cette raison, cette technique n'a pas été retenue pour l'évaluation finale.

*Roulis / Roulis.* La technique composée de roulis successifs a obtenu des résultats décevants lors des évaluation préliminaires. Les participants ont eu des difficultés à réaliser le second geste de roulis. Pour réaliser le second geste, une grande partie des participant a ressenti le besoin de revenir dans une position de repos augmentant ainsi la durée du geste. En outre, un grand nombre d'erreurs a été relevé pour les participants ne suivant pas cette stratégie.

*Boutons directionnels (BD) / BD.* Nous avons considéré cette technique comme la technique de référence dans notre évaluation pour sa simplicité et sa familiarité. C'est la technique la plus répandue sur les télécommandes actuelles. Par ailleurs, cette technique n'implique pas de mouvement de la télécommande et n'est pas affectée par la durée du geste associé.

*BD / T&L.* L'appui d'un bouton permet de choisir une direction et de délimiter le geste dans l'air. Cela permet ainsi de réduire la sélection d'une action en comparaison avec la technique précédente qui requiert deux actions distinctes pour sélectionner un élément. Dans cet espace, de nombreuses techniques illustrent cet avantage.

*BD / Roulis.* Pour des raisons similaires à la combinaison BD / T&L, il est intéressant de combiner BD avec un geste de roulis.

*T&L/ BD.* Les combinaisons inverses pourraient avoir un intérêt comparable. Cependant, contrairement aux techniques précédentes qui possèdent un délimiteur de geste implicite, celle-ci nécessite l'appui d'un bouton supplémentaire pour délimiter le geste. Le respect de l'ordre des actions pour ces techniques (*i.e.* appuyer sur la gâchette, effectuer un geste et presser sur un bouton directionnel) a entraîné trop d'erreurs lors des évaluations préliminaires.

*Roulis / BD.* Pour la même raison, nous avons éliminé la combinaison de l'appui directionnel et du geste de roulis.

En résumé, les techniques candidates combinant les trois dimensions pour évaluation sont : T&L/T&L, T&L/Roulis, BD/BD, BD/T&L, BD/Roulis (Figure 5.10).

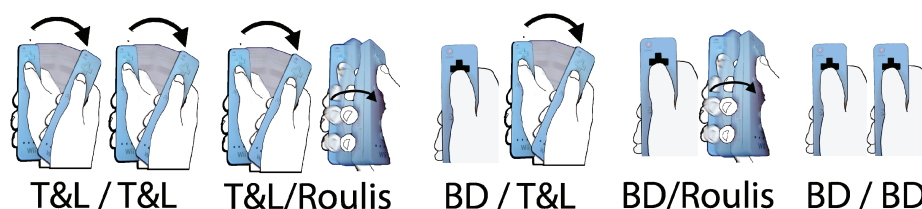


FIGURE 5.10: Techniques retenues pour l'évaluation

## 5.7 EVALUATION DES TECHNIQUES

### 5.7.1 Expérimentation

L'intérêt de cette expérimentation est de comparer l'efficacité des cinq techniques retenues précédemment dans une situation où le menu n'est pas affiché à l'écran, représentant l'interaction par un utilisateur expert. Comme ex-



pliqué précédemment, la technique BD/BD est notre technique de référence.

La tâche de cette expérimentation consiste à sélectionner un élément du sous-menu d'un menu à deux niveaux privé de son affichage. Pour simuler cette tâche et évaluer la performance des techniques, nous avons demandé aux participants de reproduire une séquence de gestes correspondant aux combinaisons définies précédemment. Pour chaque geste, nous avons défini des stimuli graphiques illustrés en figure 5.11. Pour chaque essai, deux symboles étaient affichés au centre de l'écran. Ce couple de symbole indique la première et la seconde modalité. Un exemple pour la technique BD/Roulis est illustré à la figure 5.12. Les stimuli, différents pour chaque modalité, accentuent la correspondance avec le geste et permettent d'éviter toute source de confusion. Une fois le geste effectué par le participant, le système affichait le geste reconnu à l'écran pendant 3 secondes. Ce retour visuel avait pour but l'amélioration de la réalisation des gestes par les participants et par la même occasion de présenter un retour après la sélection d'une commande comme dans une situation en contexte réel. De plus, le délai d'affichage du retour visuel permettait aux participants de revenir en position de repos avant l'essai suivant.

Le même équipement et les mêmes logiciels utilisés dans les expérimentations précédentes nous ont servi à récolter les données. Comme expliqué auparavant, les accéléromètres seuls ne sont pas capables de détecter une rotation de lacet. Pour cette raison, nous avons équipé la *Wiimote* d'une extension *Wii Motion Plus*. Le *Wii Motion Plus* permet de compléter les accéléromètres intégrés dans la *Wiimote* par des capteurs gyroscopiques. Le nombre de bouton directionnels de la *Wiimote* étant limitée à quatre comme la plupart des télécommandes disponibles sur le marché, le nombre d'éléments sélectionnables dans le menu a été limité à quatre éléments. De même, nous n'avons considéré que les quatre directions cardinales pour les gestes T&L et quatre distances angulaires pour les gestes de roulis :  $90^\circ$ ,  $45^\circ$ ,  $-45^\circ$ , et  $-90^\circ$ , la valeur numérique donnant l'amplitude et le signe la direction du geste. Ceci cor-

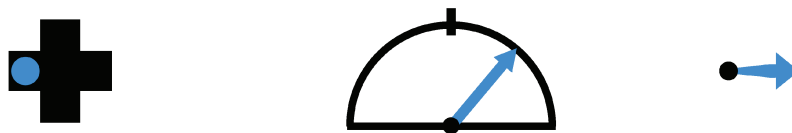


FIGURE 5.11: Stimulus graphique pour chaque modalité : à gauche bouton directionnel, au centre roulis, et à droite tangage et lacet.

respond à une combinaison de deux menus comprenant chacun 4 éléments comme illustré à la Figure 5.9.

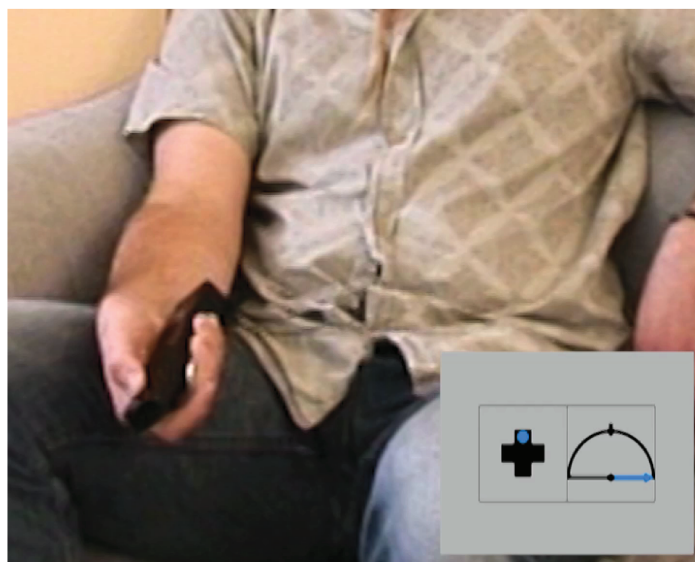


FIGURE 5.12: Un utilisateur réalisant un geste de BD/Roulis pendant l'expérimentation. L'image du participant est en miroir car filmée par une caméra.

Avant de débiter la session d'évaluation, les participants ont eu l'opportunité de pratiquer les différentes techniques pendant trois séries de seize sélections aux maximum. L'expérimentation comptait cinq séries de seize éléments qui apparaissaient dans un ordre aléatoire à l'écran. L'ordre des techniques a été compensé parmi les participants à l'aide d'une distribution en carré latin. Treize participants droitiers, dont sept personnes de sexe masculin, ont pris part à cette expérimentation de 45 minutes. Lors de l'expérimentation, nous avons mesuré le temps de sélection et le nombre d'erreurs de sélection. En tout, l'expérimentation a compris :  $13 \text{ utilisateurs} * 5 \text{ techniques} * 5 \text{ séries} * 16 \text{ éléments}$ , soit 5200 sélections.

À la fin de l'expérimentation, les participants ont répondu à un questionnaire pour évaluer les préférences subjectives sur des critères tels que la vitesse, la complexité, et l'amusement liés à la pratique des différentes techniques.

### 5.7.2 Résultats et discussion

Une analyse de la variance a révélé un effet significatif des techniques sur le temps de sélection ( $F_{4,48} = 25.34, p < 0,0001$ ). Le test de comparaison multiple de Tukey indique que les techniques BD/T&L (1,39 secondes) et BD/BD (1,40s) sont significativement plus rapide que les techniques BD/-Roulis (1,65s) et T&L/T&L (1,68s), elles même significativement plus rapide que la technique T&L/Roulis (2,02s). Les résultats sont résumés dans la figure 5.13

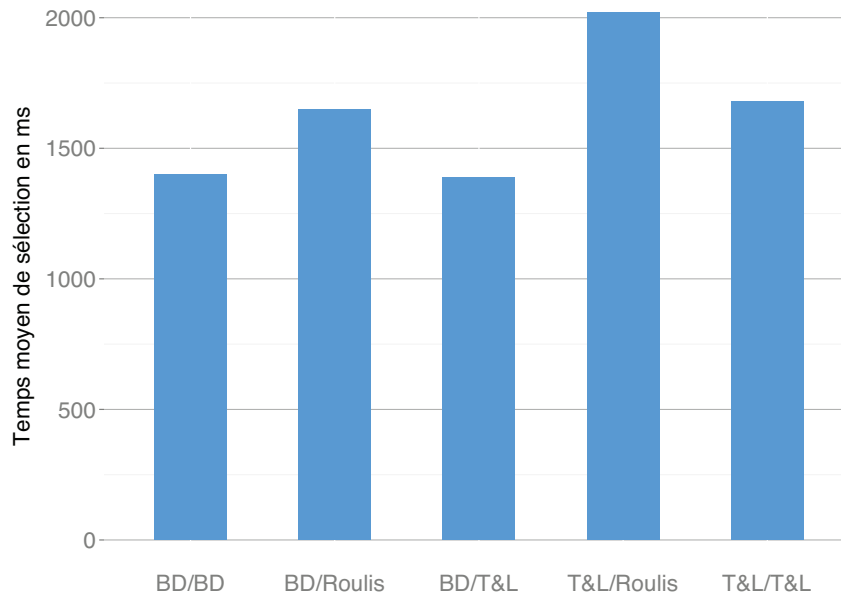


FIGURE 5.13: Temps moyen de sélection pour chaque combinaison

En terme d'erreurs de sélection, un effet significatif des techniques a été relevé ( $F_{4,48} = 17,52, p < 0,0001$ ). Les résultats sont résumés dans la figure 5.14. Le test de comparaison multiple de Tukey a montré que les techniques BD/T&L (98,4%) et BD/BD (97,4%) étaient significativement plus précises que les techniques T&L/T&L (89,5%) et T&L/Roulis (86,8%). La technique BD/Roulis (94,2%) était aussi significativement plus précise que la technique T&L/Roulis. Les raisons de l'imprécision de certaines techniques est liée d'une part aux erreurs de sélection, mais d'autre part à l'impossibilité des participants à réaliser une sélection à cause de gestes trop petits ou la réalisation de translations à la place de rotations. Cette observation est particulièrement

vraie pour les techniques T&L/T&L et T&L/Roulis (avec 2,5% d'imprécision sur l'ensemble total des sélections pour T&L/T&L et 1,3% pour T&L/Roulis).

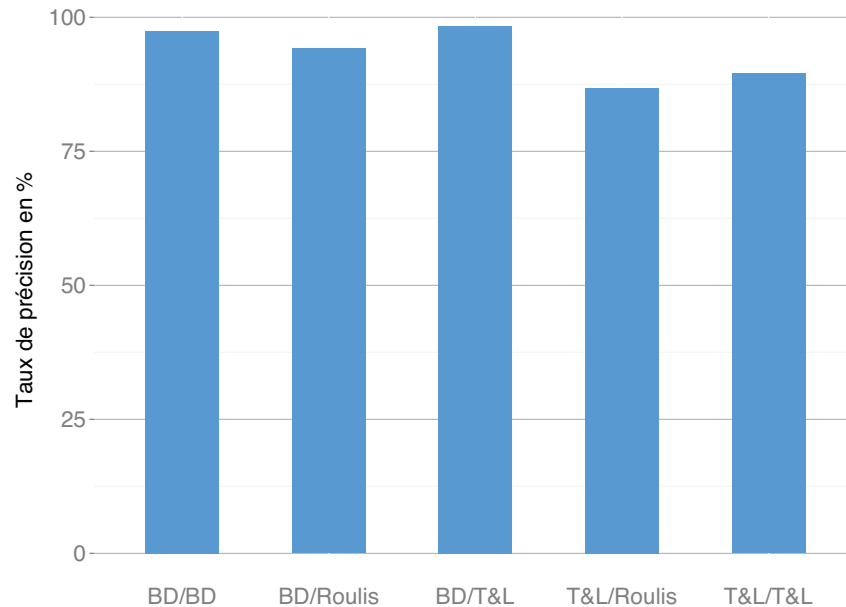


FIGURE 5.14: Taux moyen de précision pour chaque combinaison

Les résultats des questionnaires post-expérimentaux indiquent une préférence générale pour la technique T&L/Roulis avec une moyenne de 4,8, vient ensuite BD/Roulis avec 3,5, puis T&L/T&L avec 2,8, suivi de BD/T&L avec 2,4 et enfin BD/BD avec 1,5. Nous avons par ailleurs recueilli les évaluations subjectives des 5 techniques illustrées par le tableau 5.4. L'analyse avec le test de Kruskal-Wallis soulève la perception importante de rapidité, de précision, de simplicité pour la technique BD/BD. Notons aussi que cette technique est perçue comme la moins fatigante de toutes. La technique BD/T&L a été perçue plus rapide que T&L/Roulis et T&L/T&L. Elle a été également perçue comme plus simple que T&L/Roulis. Enfin, aucune technique n'a été significativement jugée comme plus amusante qu'une autre.

La stratégie de combiner les boutons et l'interaction gestuelle s'est avérée remarquablement efficace pour sélectionner des commandes. Les techniques BD/BD, BD/T&L, et dans une moindre mesure BD/Roulis (resp. 1.40s, 1.39s et 1.65s) ont été les techniques les plus rapides. Ces techniques ont aussi exhibé d'importants taux de précision (resp. 97.4%, 98.4% and 94.2%). De prime

		Technique				
		BD/BD	BD/T&L	BD/Roulis	T&L/T&L	T&L/Roulis
Caractéristique	Rapide	6,4	5,5	4,5	4,8	3,1
	Précis	6,8	5,4	4,5	4,9	3,8
	Agréable	6,5	5,5	4,1	5,5	3,3
	Fatigant	1,2	2,6	4,5	3,2	5,4
	Facile à apprendre	6,9	5,9	4,8	5,8	3,6
	Amusant	4,1	5,1	4,3	5,2	4,1

TABLE 5.4: Préférences subjectives moyennes sur une échelle de 1 à 7.

abord, on aurait pu penser que la technique BD/BD serait la plus performante grâce à la rapidité des deux pressions successives sur des boutons idéalement disposés sur la télécommande. A contrario on pouvait craindre à ce que les gestes dans l'espace demandent plus de temps et entraînent d'avantage d'erreurs. Cependant nos résultats indiquent une performance quasi-similaire entre les techniques BD/BD et BD/T&L. De plus, malgré des performances légèrement en deçà des meilleures techniques, notons quand même la très bonne performance de la technique BD/Roulis.

Ces résultats soulignent l'efficacité de combiner boutons et gestes. En outre, ces techniques sont compatibles entre elles. Toutes ces techniques peuvent cohabiter et augmenter ainsi le nombre de commandes disponibles. L'ajout de techniques d'interaction gestuelle aux télécommandes comprenant déjà des boutons permet d'augmenter leur expressivité à peu de frais et sans remettre en cause les usages habituels. Par exemple il est possible d'activer des quasimodes [Raskin, 2000] : appuyer sur un bouton sans mouvement particulier de la télécommande produirait l'action habituelle quand celui-ci est relâché, alors qu'avec un geste de T&L ou un geste de roulis pourrait déclencher une action spécifique.

Les techniques entièrement gestuelles ne sont pas toutes prometteuses. La technique T&L/T&L a donné de bons résultats contrairement à la technique T&L/Roulis dont les résultats sont un peu décevants.

Une première raison d'un tel écart de performance peut être expliqué par la quantité d'attention nécessaire pour combiner les gestes de T&L avec les gestes de roulis. Des contraintes biomécaniques pourraient être une seconde raison. Les participants ont exprimé la difficulté d'effectuer un mouvement de roulis pour plus de trois éléments après avoir effectué un geste de T&L :

« La technique de T&L/Roulis pourrait être ma technique préférée si il n'avait seulement que trois variations angulaires ».

Certaines combinaisons de gestes ont vraisemblablement été pénalisées par une précision insuffisante du système de reconnaissance. L'analyse des enregistrements vidéos a en effet montré qu'un nombre important de gestes correctement effectués par les utilisateurs n'ont pas été reconnus (environ 50% pour T&L/T&L et 17% pour T&L/Roulis). La télécommande utilisée est équipée d'un gyroscope dont la mesure dérive continuellement et augmente avec le temps et les mouvements [Wingrave *et al.*, 2010]. Il est nécessaire de recalibrer le capteur régulièrement ou de compenser cette dérive. De plus, un traitement sur le signal est indispensable pour éliminer le bruit et fusionner correctement les mesures angulaires fournies par les capteurs. Malheureusement, l'algorithme utilisé dans cette étude n'était sans doute pas assez sophistiqué. Un algorithme de traitement de signal plus performant tel que les filtres de Kalman [Kalman, 1960] ou le *1 $\epsilon$  Filter* [Casiez *et al.*, 2012] pourraient lisser le signal et améliorer la reconnaissance et, de ce fait, les résultats. Nos résultats peuvent donc être considérées comme des valeurs limites minimales.

### 5.7.3 Mode débutant du menu

Le mode novice (Figure 5.15) présente un menu à deux niveaux avec une disposition des éléments dépendant de la technique d'interaction utilisée. Selon la technique, les éléments de premier niveau et de second niveau sont sélectionnés soit par BD ou soit par des gestes de T&L. Les gestes de roulis permettent également de sélectionner des éléments du second niveau. Le mode novice est activé grâce à une pression d'un bouton pendant 300 millisecondes et en l'absence de tout geste interprétable par le système.

Chaque élément du premier niveau du menu représenté sous une forme circulaire permet d'accéder à un sous menu composé de 8 éléments également disposés circulairement sauf pour les gestes de roulis. Dans ce cas, la disposition est semi-circulaire suggérant ainsi l'utilisation de ce type de geste. Pour des raisons techniques, seulement 4 éléments situés sur les directions cardinales sont utilisés dans le cas où la technique implique BD ou si les gestes diagonaux pour la technique T&L sont désactivés. Les éléments sont sélectionnés de la même manière en mode débutant et en mode expert. Le menu semi-circulaire fonctionne selon le même principe mais seulement avec 5 éléments (Figure 5.15, droite).

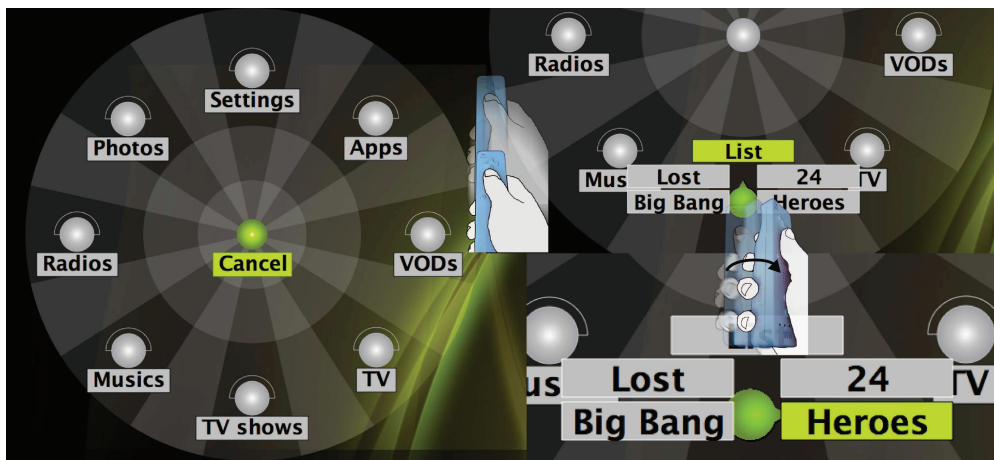


FIGURE 5.15: Menu correspondant à une interaction T&L/Roulis. Le menu pour le roulis est un menu semi-circulaire de 5 éléments.

Pour palier à la limitation du nombre d'élément disponibles dans les menus circulaires, nous avons aussi envisagé le scénario dans lequel activer un élément du menu de premier niveau n'ouvre pas un sous-menu mais une représentation tabulaire contenant autant d'éléments que nécessaires. Par exemple, sélectionner l'élément "TV Shows" dans le menu entraînerait l'affichage du tableau illustré en Figure 5.16. L'utilisation des boutons directionnels permet alors de naviguer à l'intérieur de la table tandis que les gestes de tangage et lacet servent à passer à la table précédente ou la suivante.

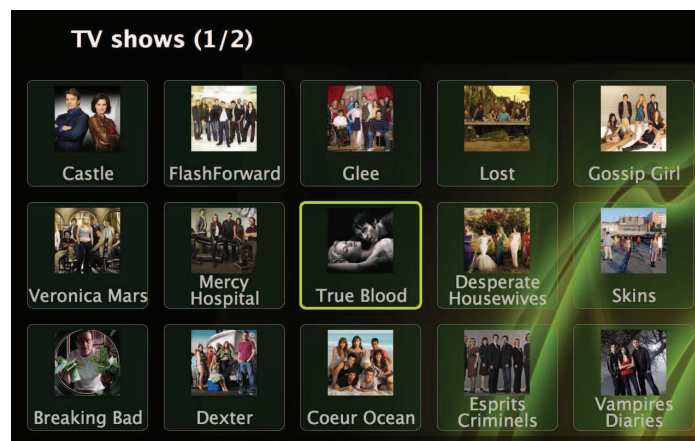


FIGURE 5.16: Représentation tabulaire des éléments.

Similairement aux *Marking Menus*, les utilisateurs peuvent facilement apprendre le fonctionnement de la technique d'interaction grâce à la correspon-



dance directe entre l'orientation de l'élément affiché et le geste que l'utilisateur doit réaliser avec le pouce ou le poignet. Un point crucial dans la technique est qu'elle encourage une transition fluide et implicite entre l'usage en mode novice et l'usage en mode expert étant donné que l'utilisateur effectue les mêmes gestes dans les deux modes [Kurtenbach et Buxton, 1991].

#### 5.7.4 Perspectives

Bien que la première étude sur la correspondance entre les gestes et le mouvement du dispositif ait montré une préférence générale pour les rotations, quelques utilisateurs ont spontanément utilisé des translations. Suite à ce constat, notre travail exploratoire s'est essentiellement focalisé sur l'étude de geste basé sur la rotation du dispositif. Cependant tenir compte des translations du dispositif pourrait encore augmenter l'expressivité de la télécommande grâce, par exemple, à la localisation d'un groupe de commandes dans un espace 3D. C'est le cas de *Cube Key* qui utilise les translations en 3 dimension de la télécommande pour sélectionner un caractère dans une matrice cubique pour de la saisie de texte sur un écran distant [Shoemaker et al., 2010].

Étant donné les limitations techniques de la *Wii mote* qui n'offre qu'un pavé directionnel à quatre directions, nous avons seulement considéré quatre éléments par niveau de menu. Jones et al. [2010] rapportent la difficulté à effectuer des gestes diagonaux. Pourtant, les résultats de notre première expérimentation indiquent que les utilisateurs réalisent naturellement des gestes diagonaux en combinant tangage et lacet lorsque le dispositif d'affichage est déporté. Dans leur étude, la *Wii mote* n'était pas équipée de gyroscopes. Un approfondissement est nécessaire pour évaluer la réalisation de gestes diagonaux et, par la même occasion, l'efficacité de notre vocabulaire gestuel dans le cas où le menu comporte 8 éléments par niveau.

Le mode expert nécessite un appel à la mémoire pour effectuer le geste associé à la commande désirée. Une évaluation exhaustive de ce processus implique une phase de mémorisation et une phase de rappel [Kurtenbach et al., 1993; Bailly et al., 2008] nécessitant une expérimentation de longue durée, irréalisable avec le nombre conséquent de techniques évaluées dans notre étude. Contrairement aux autres techniques qui partagent des propriétés similaires aux *Marking Menus*, les gestes basés sur le roulis ne proposent pas une sélection directionnelle mais une sélection basée sur une distance angulaire à parcourir. La mémorisation des commandes entre ces deux types de

de sélection mériterait aussi une recherche plus approfondie.

## 5.8 CONCLUSION

Dans cette partie, nous avons investigué les nouvelles possibilités offertes par les gestes 3D dans l'air pour augmenter et améliorer l'interaction avec les télécommandes traditionnelles. Les télécommandes tendent à être surpeuplées de boutons, une conséquence directe d'un nombre toujours grandissant de nouvelles fonctionnalités dans les systèmes de télévision interactive. Les gestes en l'air semblent être appropriés pour limiter ce problème et offrir la capacité d'atteindre les commandes ou les contenus favoris sans avoir à regarder la télécommande.

La première étude a souligné l'influence de la localisation de l'écran d'affichage sur les types de mouvement du dispositif pour une tâche de sélection de direction. Particulièrement, nous avons montré la pertinence de l'usage des rotation de tangage et lacet pour l'interaction avec une télécommande et un écran distant.

La seconde étude a montré la capacité des utilisateurs à utiliser un geste de roulis pour sélectionner précisément 5 à 7 éléments avec un algorithme d'apprentissage.

En nous appuyant sur ces résultats, nous avons proposé 5 nouvelles techniques d'interaction tirant parti des avantages des différentes dimensions (boutons et gestes) et leurs combinaisons. Nous avons ensuite évalué leur performance. Les résultats suggèrent que les gestes en l'air sont une ressource d'entrée additionnelle prometteuse. Les techniques hybrides combinant les gestes en l'air et les gestes sur la surface de la télécommande se sont révélés particulièrement efficaces avec l'avantage notable d'être compatible avec les techniques d'interaction basées sur les boutons déjà existants sur ces systèmes. Par ailleurs, notre adaptation des *Marking Menus* 2D pour les gestes en 3D en utilisant les boutons directionnels et les rotations de tangage et lacet est adéquate pour l'interaction en mode expert et sans affichage des menus à l'écran. Enfin, les techniques d'interaction purement gestuelles présentées dans cette partie ont démontré des performances légèrement inférieures mais en partie liées aux limitations techniques.



---

## BELLY GESTURES : INTERACTION GESTUELLE NON INSTRUMENTÉE

---

### 6.1 INTRODUCTION

Dans la partie précédente, nous avons montré comment en augmentant les capacités de la télécommande traditionnelle, il est possible d'améliorer les performances des téléspectateurs pour sélectionner des commandes fréquentes ou des contenus favoris. Bien que la télécommande matérielle soit un objet communément accepté dans l'environnement du salon, son usage présente quelques inconvénients. D'une part, comme tout dispositif mobile, la télécommande est dépendante de la source d'énergie par laquelle elle est alimentée [Villar et Hodges, 2010]. De plus, la télécommande est aussi victime de mal-adresses de téléspectateurs peu attentifs dans un environnement animé. Par exemple, une chute malencontreuse ou une exposition à des liquides peut la rendre inutilisable. Enfin, et surtout, cet objet est facilement égaré [Freeman T. et Craig D., 1995] et il est fréquemment hors de portée de l'utilisateur qui doit alors se déplacer pour l'atteindre.

L'utilisation de dispositifs plus élaborés, comme le smartphone ou la tablette tactile, est de plus en plus répandue dans l'environnement domestique. Ces dispositifs permettent d'enrichir l'expérience télévisuelle et de compléter le contenu diffusé sur la télévision. Cependant, du seul fait de leur vocation multi-fonctions, ils sont d'un emploi malcommode quand il s'agit d'exécuter rapidement une commande isolée, par exemple allumer, éteindre le téléviseur ou changer de chaîne.

Une solution alternative à la télécommande physique est l'utilisation de gestes dans l'espace. Comme nous l'avons vu dans l'état de l'art, ils peuvent être à la longue fatigants et imprécis [Baudel et Beaudouin-Lafon, 1993]. Une autre solution consiste à utiliser le corps comme une surface d'interaction. Plusieurs travaux ont examiné cette possibilité dans des contextes de mobilité. Mais, à notre connaissance, aucune recherche n'a été conduite sur ce

sujet dans le contexte de la télévision interactive. Dans la suite de ce chapitre, nous présentons un bref aperçu de la littérature sur ce thème, suite à quoi, nous proposons l'usage de l'abdomen comme surface interactive et montrons comment il est possible de réaliser des gestes pour exécuter des commandes rapidement.

## 6.2 BELLY GESTURES : EXPLORATION DES GESTES SUR LA SURFACE DE L'ABDOMEN

Les travaux de recherche sur l'interaction sur la surface du corps ne cessent de fleurir. De nombreux efforts visent à concevoir de nouvelles technologies capables de capturer les contacts et de les localiser sur le corps [Harrison *et al.*, 2010, 2011; Nakatsuma *et al.*, 2011; Holz *et al.*, 2012; Sato *et al.*, 2012a], en adaptant parfois des techniques d'interaction déjà existantes et répandues sur d'autres support tels que les tables ou tablettes interactives [Harrison *et al.*, 2011]. Étonnamment, une attention minime a été accordée à la compréhension des facteurs humains influençant l'interaction sur la surface des parties du corps [Gustafson *et al.*, 2011; Lin *et al.*, 2011].

### 6.2.1 Les gestes sur la surface du corps

Dans cette partie, nous présentons ces différents types de gestes au travers de la récente littérature sur ce sujet.

#### 6.2.1.1 Les gestes de "tap"

Des études ont été menées pour mesurer la capacité des utilisateurs à effectuer une sélection dans un vocabulaire de commandes au moyen de brèves pressions à différents endroits du corps. Par exemple, Lin *et al.* [2011] rapportent que la plupart des utilisateurs sont capables de diviser la surface de leur avant-bras en six zones distinctes sans engager leur attention visuelle sur cette partie du corps. Les auteurs expliquent que les positions mentales perçues lors du contact du doigt avec l'avant-bras diffèrent des mesures réelles et que les utilisateurs possèdent une représentation intrinsèque de l'espace de cette surface lorsque le nombre de zone à discriminer devient important. En pressant avec la main sur une partie de l'avant-bras opposé, Shoesense [Bailly *et al.*, 2012] reconnaît la forme dessinée par les deux bras pour exécuter la commande associée (Figure 6.4).

*PinchWatch* donne accès aux commandes par des pressions succinctes entre le pouce et les autres doigts de la main [Loclair *et al.*, 2010]. En s'appuyant sur

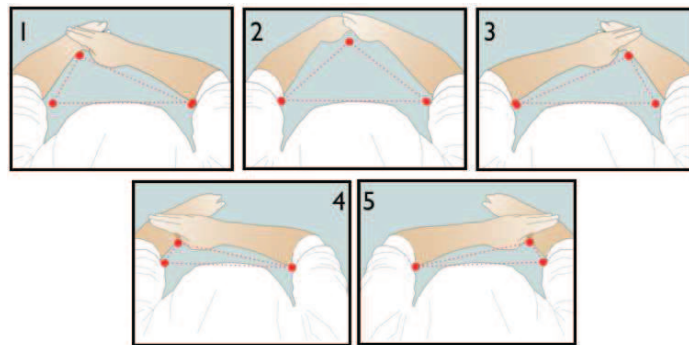


FIGURE 6.1: *Shoeseense* : un ensemble de gestes triangulaires [Bailly *et al.*, 2012].

la mémoire spatiale des utilisateurs, *Imaginary Phone* [Gustafson *et al.*, 2011] tire avantage de la mémorisation de la position des icônes des applications d'un smartphone pour en transférer l'interface sur la main (Figure 6.2). Les auteurs ont évalué la précision du pointage sur la paume de la main et ont trouvé que leur technique était suffisamment précise pour contrôler les applications sur les dispositifs mobiles. Les points de repères visuels facilitent la sélection de la position souhaitée mais la technique monopolise l'attention visuelle. En utilisant la pression localisée, seulement un petit ensemble de commandes peuvent être assignées. Par ailleurs, cette modalité d'interaction est fortement sujette aux activations involontaires.



FIGURE 6.2: *Imaginary Phone* : l'utilisateur utilise la paume de sa main comme surface de substitution à son téléphone. Grâce à la mémoire spatiale, l'utilisateur exécute une commande sur son téléphone en touchant la paume de sa main [Gustafson *et al.*, 2011].

#### 6.2.1.2 *Les gestes de glissement*

Les gestes de glissement sont habituellement adéquats pour contrôler des paramètres continus. Glisser l'extrémité du pouce le long des doigts de la main [Loclair *et al.*, 2010] ou la main sur le bras opposé [Bailly *et al.*, 2012] permet par exemple d'ajuster le volume sonore. De même, faire glisser son pouce circulairement sur la paume de sa main [Loclair *et al.*, 2010] permet de naviguer dans une chanson. De tels gestes peuvent aussi débloquent un dispositif mobile [Harrison *et al.*, 2011].

Bien que des travaux se soient intéressés au geste de glissement sur les surfaces rigides planes [Baglioni *et al.*, 2011] et courbes [Voelker *et al.*, 2012], peu d'attention a été portée au cas d'une surface élastique et souple comme la peau recouvrant les muscles du corps. Voelker *et al.* [2012] rapportent que la précision du glissement dépend de la précision avec laquelle l'utilisateur est capable d'identifier la direction correcte, même avec l'aide de la vision. Ce point paraît important tant les parties du corps sont plus ou moins mobiles les unes par rapport aux autres et peuvent revêtir différentes orientations. Quand la technique implique les membres supérieurs, ce qui est le cas le plus fréquent dans la littérature en Interaction Homme-Machine, on peut se demander si leur orientation, par nature très variable, ne serait pas à même d'affecter l'efficacité ou la stabilité des représentations mentales.

#### 6.2.1.3 *Le geste sémiotique*

En utilisant les capacités électro-conductrices du corps humain, [Sato *et al.*, 2012a] réussissent à reconnaître les gestes sémiotiques (cf la section 4.2.2 de ce manuscrit) comme la saisie d'un objet ou la couverture de ses propres oreilles avec ses mains. Porteurs de signification, les gestes de cette catégorie pourraient en principe être utilisés assez naturellement comme raccourcis vers des commandes sémantiquement correspondantes.

#### 6.2.2 *Localisation des gestes sur le corps*

Dans la brève revue de la littérature qui suit, nous nous contenterons d'une taxonomie élémentaire du corps humain distinguant simplement quatre membres montés sur un tronc, la région abdominale de ce dernier retenant spécialement notre attention. La question de la localisation corporelle idéale de l'interaction sans dispositif physique occupe de nombreux chercheurs.



### 6.2.2.1 Les gestes sur les bras

De nombreuses techniques d'interaction utilisent la surface du bras [Harrison *et al.*, 2010; Lin *et al.*, 2011; Bailly *et al.*, 2012]. Cette partie du corps offre d'intéressantes caractéristiques telles que des propriétés bio-acoustiques grâce à la propagation du son le long de la peau [Harrison *et al.*, 2010; Lin *et al.*, 2011]. Elle permet aussi un accès aisé avec la main opposée pour effectuer des gestes discrets qui demandent peu d'effort [Bailly *et al.*, 2012]. Enfin, elle procure une bonne résistance aux dommages infligés par chocs, pour les dispositifs portés ou implantés sous la peau [Holz *et al.*, 2012] (Figure 6.3). Cette dernière approche possède la particularité de rendre le dispositif d'interaction complètement invisible, continuellement accessible et non intrusif dans les activités quotidiennes des utilisateurs. On observera que ces techniques ne permettent pas l'interaction bi-manuelle au sens où l'un des deux membres est mobilisé comme cible passive du geste.

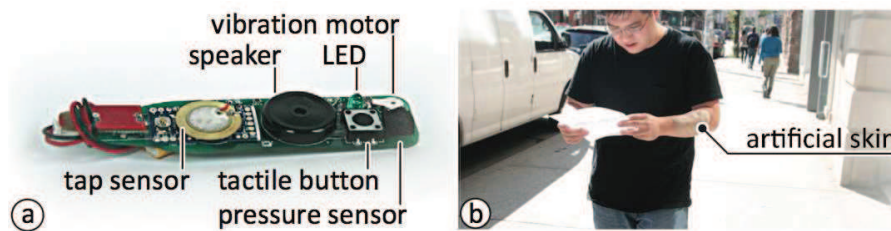


FIGURE 6.3: *Implanted user interfaces* : des capteurs sont intégrés sous une peau artificielle pour par exemple détecter une pression [Holz *et al.*, 2012].

### 6.2.2.2 Les gestes sur les mains

La main fournit une interaction précise lors de la manipulation des outils du quotidien. Cette observation a motivé de nombreux chercheurs dans l'évaluation de nouvelles techniques de capture pour interagir avec cette partie du corps [Loclair *et al.*, 2010; Sato *et al.*, 2012a; Gustafson *et al.*, 2011]. Alors que *PinchWatch* [Loclair *et al.*, 2010] est capable de détecter les gestes unimanuels tels que les pincements entre le pouce et un autre doigt, des glissements ou des gestes simples des doigts sur la paume, *Touché* [Sato *et al.*, 2012a] reconnaît des interactions bimanuelles comme agripper son poignet, ou encore le contact doigts à doigts entre les deux mains (Figure 6.4). La main possède aussi de nombreux repères visuels. Comme on l'a vu, *Imaginary Phone* tire parti des repères remarquables de la main pour transférer l'interface d'un téléphone mobile sur cette partie du corps.

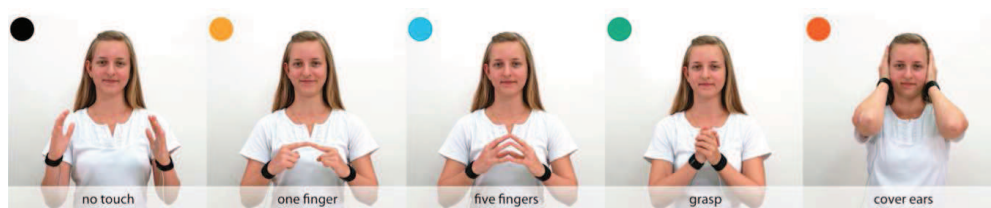


FIGURE 6.4: *Touché* : états détectés avec des gestes sur le corps et des capteurs portés sur les poignets.

Les mains, qui évidemment fournissent de nombreux avantages comme agents et cibles potentiels de l'interaction, sont malheureusement souvent déjà engagées et leur disponibilité ne va pas de soi.

#### 6.2.2.3 Les gestes sur les autres parties du corps

D'autres travaux se sont intéressés à l'exploration d'autres parties du corps. [Gemperle et al. \[1998\]](#) fournissent des recommandations pour la conception des systèmes portés. Ils suggèrent que l'endroit idéal pour un tel dispositif devrait être une partie du corps qui minimise les mouvements et la flexibilité des membres, particulièrement quand le corps est en déplacement, et qui maximise la surface disponible. Il est intéressant de noter que [Karrer et al. \[2011\]](#) ont conduit une étude pour déterminer le meilleur endroit du corps pour interagir avec *Pinstripe*. *Pinstripe* est une technique d'interaction basée sur le pincement des plis d'un vêtement. En faisant rouler le pli entre deux doigts, cette technique permet de contrôler finement une variable continue sans retour visuel. Il est alors par exemple possible de contrôler du défilement de liste. Après avoir réalisé la technique d'interaction sur différentes parties de leur corps, les participants ont apprécié le confort d'utilisation des endroits pour la réalisation du geste de pincement et roulement. Les résultats indiquent que les bras, les avant-bras et le sternum sont les parties du corps les plus commodes pour interagir avec cette technique. Les hanches, la taille et les cuisses, bien que jugées moins commodes pour cette technique, n'ont pas pour autant été écartées par les participants comme possibles supports d'interaction. [Wagner et al. \[2013\]](#) rapportent que, dans une tâche d'acquisition de cibles sur un écran distant et sur le corps, les utilisateurs ont préféré atteindre les cibles avec la main sur la partie haute du corps (des cuisses aux épaules) avec une légère préférence pour le torse plutôt que le bras dominant.

Pour de nombreuses raisons, les mains et les bras ont mobilisé l'attention dans la recherche sur l'interaction corporelle. Néanmoins, les potentialités offertes par le tronc, au niveau du torse ou de l'abdomen, ont à peine commencé

à être explorées. Or les perspectives sont prometteuses.

### 6.2.3 *Propriétés de l'abdomen*

L'abdomen est la région du tronc située entre le thorax et le pelvis. La surface de l'abdomen possède un certain nombre de propriétés intéressantes pour notre propos. Dans cette partie, nous décrivons et illustrons ces avantages, sans ignorer les limites qui leurs sont associées.

#### 6.2.3.1 *Une surface relativement large*

Par comparaison, par exemple, avec la surface disponible sur un bras ou sur la paume d'une main, l'abdomen fournit une surface relativement large, de l'ordre de  $20 \times 20 = 400 \text{ cm}^2$ . Si un bras se prête bien à la tâche de glissement 1D [Lin et al., 2011], l'abdomen se présente d'entrée de jeu comme une surface d'interaction 2D.

#### 6.2.3.2 *Une surface accessible des deux mains*

L'abdomen est accessible pour les deux mains de manière indépendante, ouvrant la voie à la vaste famille des techniques d'interaction bimanuelles. Par exemple, il est possible d'adapter des techniques d'interaction évoluées comme le zoom et la rotation effectuées couramment avec deux doigts sur les tables et tablettes interactives. Plus généralement, il est possible de concevoir des techniques impliquant non pas seulement les deux mains, mais simultanément plusieurs doigts des deux mains. Pour ne mentionner qu'un seul exemple suggestif, les utilisateurs pourraient naviguer et sélectionner des commandes sur l'abdomen avec la technique *Finger Count* [Bailly et al., 2010].

L'interaction sur l'abdomen est aussi propice aux "*Whack Gestures*" qui nécessitent peu d'attention visuelle et un contrôle moteur grossier. Ces gestes sont généralement effectués dans des situations qui perturbent une autre activité dans laquelle l'utilisateur est déjà engagé [Hudson et al., 2010].

#### 6.2.3.3 *Une surface relativement stable*

Le tronc n'est pas exposé aux mouvements qui affectent constamment les membres, constituant de ce fait une partie stable du corps. Karrer et al. [2011] ont fait remarquer que le sternum était beaucoup moins mobile que les bras dans une expérimentation où les utilisateurs étaient invités à marcher. Ces

auteurs suggèrent que le contrôle sur textile sur le thorax pourrait être comode pendant les activités sportives. Par ailleurs, [Wagner et al. \[2013\]](#) recommandent également l'usage de cette partie du corps pour des gestes qui pourraient demander une grande précision ou de mouvements coordonnés de plusieurs parties du corps. À l'évidence l'abdomen, autre partie du tronc, constitue également une ressource pour l'interaction pendant les activités physiques.

#### 6.2.3.4 *Gestes accommodants*

L'abdomen est accessible de manière quasi-permanente. Cette partie du corps étant localisée en plein champ d'action des mains, une amplitude minimale de mouvement est nécessaire pour l'atteindre. L'abdomen constitue même l'une des positions naturelles de repos de la main. L'abdomen n'étant généralement pas impliqué dans les activités quotidiennes (à la différence des membres supérieurs), il est disponible sans coût, ce qui en fait un candidat idéal comme organe d'entrée d'information dans la plupart des situations.

#### 6.2.3.5 *Une surface offrant un repérage polaire*

Dans *Imaginary Phone*, [Gustafson et al. \[2011\]](#) tirent avantage des repères mémorables de la main pour faciliter le transfert de l'interface d'un téléphone mobile sur la main. En s'appuyant sur l'idée que ces repères permettent de créer des associations symboliques, les auteurs profitent de la ressemblance évidente entre l'écran d'accueil matriciel des téléphones mobiles et les repères visuels de la main comme la démarcation des différentes phalanges. L'abdomen possède lui aussi ses repères, avec la forme, plus ou moins saillante, des muscles abdominaux sous-jacents et surtout le nombril, toujours repérable, et dont la position centrale confère à la surface considérée une organisation radiale. Comme on le verra plus bas dans l'exposé de l'expérimentation que nous avons réalisée sur le sujet, il est particulièrement aisé de se représenter un geste à la surface de l'abdomen comme une direction dans un repère polaire plutôt que dans un repère cartésien, comme c'est généralement le cas sur les surfaces interactives.

#### 6.2.3.6 *Une surface continuellement perceptible*

L'appareil musculaire, les articulations et les récepteurs de la peau jouent un rôle substantiel dans le contrôle moteur et participent à la construction d'un modèle mental de représentation de la forme du corps. Par exemple, ces divers systèmes contribuent à manœuvrer notre chemin à travers les obstacles dans l'obscurité et à manipuler des objets hors de notre champ de vision en

fournissant des retours proprioceptifs et kinesthésiques [Proske et Gandevia, 2012]. Grâce à la kinesthésie, nous sommes constamment conscients de la position et du mouvement de nos membres, ce qui nous permet par exemple d'interagir sur la surface de notre ventre sans jeter un regard. La peau est aussi bien sûr une surface sensible. Bien que l'acuité du sens tactile varie grandement selon les parties du corps mises en jeu [Serino et Haggard, 2010], la combinaison simultanée des informations tactiles de plusieurs parties du corps peut améliorer la précision, notamment dans des contextes où l'attention visuelle est indisponible [Lin *et al.*, 2011].

#### 6.2.3.7 Une surface intime

En tant qu'événement enregistré en entrée, le contact de la peau sur la peau, procure un délimiteur de geste pour l'interaction corporelle. Cependant, le corps est sujet à d'incessants contacts plus ou moins contrôlés avec les personnes et les objets environnants. Au contraire, la région abdominale du corps reste strictement privée. Pour cette région comme pour d'autres parties du corps, le contact interpersonnel suscite des réactions émotionnelles fortes. Par exemple, être touché par une autre personne dans la région de la taille est considéré comme inapproprié et harcelant dans un environnement de travail [Gallace et Spence, 2010]. En tout cas l'abdomen, pour ce genre de raisons, est de fait nettement moins exposé aux contacts accidentels que les membres supérieurs et le reste du corps. Dans le contexte d'une étude dont l'objet est de concevoir des techniques d'interaction corporelles, cette particularité de la région abdominale lui confère un avantage intéressant.

### 6.2.4 Propriétés de l'abdomen en tant qu'une surface interactive

#### 6.2.4.1 Orientation

L'orientation de la surface conditionne l'interaction tactile. D'après Forlines *et al.* [2007], la forme du contact de la surface du doigt avec la surface interactive diffère lorsque l'orientation de la surface est horizontale ou verticale et cette orientation influence la précision du touché. Le choix de l'orientation peut être aussi essentiel selon le type d'interaction envisagé. Par exemple, Pedersen et Hornbæk [2012] ont mis en évidence que le pointage du doigt est plus bref sur une surface horizontale, tandis que le glissement du doigt est plus rapide sur une surface verticale. De plus, la précision du geste de glissement dépend de la précision à laquelle un utilisateur est capable d'en

identifier visuellement la direction [Voelker *et al.*, 2012].

L'orientation de la surface conditionne également la perception de l'orientation des utilisateurs lorsqu'ils interagissent avec. En particulier, Wobbrock *et al.* [2008] ont montré que, en interagissant sur la face arrière d'un appareil mobile tenu dans la main, les utilisateurs reproduisent des symboles alphabétiques de deux manières différentes. De plus, le choix de ces représentations est également influencé par la présence de retour visuel ou non. Ces résultats coïncident avec les travaux de Parsons et Shimojo [1987] qui étudient la perception de l'orientation spatiale sur la surface du corps dans différentes positions. Les résultats identifient différentes représentations mentales pour la tête ou les mains. Cependant, les résultats ne précisent pas de référentiel pour l'abdomen qui ne semble pas être gouverné par une représentation mentale spatiale particulière. De plus, ici, les expérimentations ont été réalisées en stimulant la peau avec un stylo en plastique par l'expérimentateur. Si les utilisateurs étaient amenés à réaliser des gestes sur cette surface, les informations motrices et sensorielles des membres actifs pourraient également influencer le choix d'une représentation spatiale.

Alors que l'orientation de la surface des membres est très variable, l'orientation de l'abdomen connaît peu d'états différents. En situation debout ou assise, il se présente comme une surface essentiellement verticale alors qu'en position couchée, il s'étend quasi-invariablement dans le plan horizontal. Il est intéressant de noter que dans la majorité des situations, les vecteurs normaux à la surface de l'abdomen pointent approximativement dans la direction du regard. Cette propriété, sur laquelle nous reviendrons dans la partie expérimentale de ce travail, pourrait avoir une influence sur la représentation interne de l'orientation spatiale des utilisateurs lors de l'interaction sur cette surface.

La perception de l'orientation spatiale, relative à un ensemble d'axes fixes définissant un cadre de référence, doit être basée sur une représentation interne de la gravité, et elle doit dépendre du plan dans lequel l'orientation est définie et de la main utilisée pour collecter les informations perceptives [Gentaz *et al.*, 2008]. La plupart des travaux précédents impliquent des plans frontaux et transverses situés à distance des participants. À notre connaissance aucune étude ne s'est intéressée au cas où les participants interagissent dans un plan frontal confondu avec la surface de leur propre corps. La perception de l'orientation spatiale dans cette configuration particulière reste inexplorée.

#### 6.2.4.2 *Surface non strictement planaire*

Contrairement à de nombreuses surfaces interactives comme les tables interactives, les tablettes, les téléphones mobiles équipés d'écran tactile ou encore les tableaux interactifs, l'espace abdominal n'est pas rigoureusement plat. De nombreux résultats montrent que l'interaction sur les surfaces courbes diffère de l'interaction sur les surfaces planes. Par exemple, l'inclinaison et le degré de courbure d'une surface interactive non plane influencent la précision du pointage avec le doigt et la performance dépend de la posture du doigt [Roudaut *et al.*, 2011]. Dans une autre étude, Voelker *et al.* [2012] rapportent que les gestes de glissements sont moins performants lorsque qu'ils sont réalisés à proximité d'une surface courbe à cause d'un biais de perception lié à ce type de surface. Les performances et le comportement des utilisateurs diffèrent lors d'une interaction avec ce genre de surface. Les surfaces flexibles devenant en ce moment de plus en plus accessibles et populaires, il semble important de comprendre plus en détail l'interaction sur surface courbe.

#### 6.2.4.3 *Surface élastique*

Une propriété intéressante de la peau est son élasticité, spécialement marquée dans les zones corporelles où elle recouvre du tissu musculaire. En combinaison avec celle des muscles, cette élasticité se traduit dans le fait que la surface est réversiblement déformable. En appliquant des pressions sur la peau, les utilisateurs ont la possibilité de créer des concavités locales qui pourraient être utilisées comme des puits physiques pour guider des micro-gestes tels que les *MicroRolls* [Roudaut *et al.*, 2009]. Le degré de pression pourrait aussi être utilisé comme un sélecteur de mode permettant de créer des gestes sémantiques sur la peau déjà exploré dans le cadre de surfaces rigides [Ramos et Balakrishnan, 2007; Heo et Lee, 2012]. L'utilisation d'une surface active, comme la surface du bout du doigt ou de la paume de la main et d'une surface passive support du geste, comme la surface de l'abdomen, pourrait augmenter l'échelle des niveaux de pression discrétisables par les utilisateurs. Enfin, la rigidité de cette surface peut être variable par la contraction des muscles. Il devrait être alors possible de profiter d'une surface à peu près rigide [Sato *et al.*, 2012b] et d'une surface à retour de force [Lopes *et al.*, 2013].

#### 6.2.4.4 *Limitations de la surface*

Si l'interaction sur l'abdomen semble présenter de nombreuses caractéristiques prometteuses, elle amène également certaines contraintes, qu'il convient



de comprendre pour s'en accommoder.

Du fait de la position de l'abdomen par rapport à la tête, vérifier visuellement un geste que l'on est en train d'effectuer sur cette partie du corps demande un effort spécial. Pour tirer le meilleur parti de cette contrainte, les techniques d'interaction sur cette partie du corps devraient tout simplement se passer du contrôle visuel en exploitant les retours proprioceptifs originaires, de l'effecteur actif d'une part et, d'autre part, de la surface abdominale stimulée par les contacts.

Certaines situations sont susceptibles de compromettre l'usage de l'abdomen, notamment la grossesse et l'embonpoint. Il reste qu'une meilleure compréhension du problème pourra éventuellement garantir la maîtrise des techniques d'interaction considérées quelle que soit la courbure de la surface abdominale.

L'interaction sur l'abdomen ouvre une voie de recherche encore peu explorée à ce jour. Dans la suite de ce chapitre, nous nous intéressons à la représentation mentale de gestes effectués sur la surface de l'abdomen, en nous préoccupant de la probable incidence de cette représentation sur le choix et l'exécution des gestes.

#### 6.2.5 *Expérimentation : principe de l'expérience*

Dans cette étude nous nous sommes demandés comment les utilisateurs se représentaient mentalement dans l'espace la surface de leur abdomen, sachant qu'ils devaient tracer des formes sur cette surface en continuant de regarder devant eux. Cette expérimentation s'inspire des travaux de [Parsons et Shimojo, 1987] et [Wobbrock *et al.*, 2008]. Pour comprendre le problème, considérons la Figure 6.5. On suppose qu'un participant reçoit, sur un écran placé devant lui, la consigne (délibérément peu précise) de tracer sur son abdomen le chiffre 2. L'une des solutions possibles pour le participant est de produire sur son abdomen le tracé illustré dans la Figure 6.5 à l'identique. Ce tracé aurait alors la propriété d'être lisible pour un observateur situé derrière le participant et possédant le pouvoir de lire à travers le thorax du participant. Dans un tel cas, la correspondance entre le stimulus de l'écran et le tracé abdominal est directe.

Mais en réalité le participant dispose de quatre options, a priori toutes correctes, indiquées dans la Figure 6.6, et qui se distinguent par l'existence

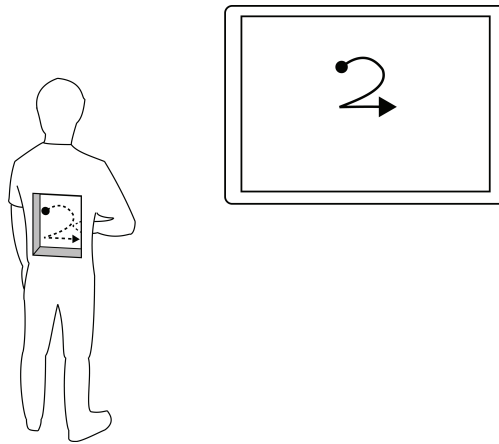


FIGURE 6.5: Point de vue de la représentation de l'orientation spatiale

ou non d'une inversion en miroir d'une part verticalement et d'autre part horizontalement. Toute la question est de savoir comment le participant "regarde" mentalement une surface de son corps qu'il perçoit uniquement grâce à la proprioception. Si, par exemple, il considère son ventre de haut en bas (en s'imaginant baisser simplement le regard), selon toute vraisemblance il dessinera son "2" avec une inversion par rapport à l'axe horizontal et sans inversion par rapport à l'axe vertical (Figure 6.6 - case sud-ouest), puisque c'est la seule option qui, depuis ce point de vue, conserve la lisibilité du chiffre.

Pour faciliter l'analyse de cette question, nous avons décidé de demander à nos participants de tracer non seulement des chiffres, qui requièrent un geste complexe d'écriture, mais aussi de simples directions linéaires identifiables par leur seule direction (haut, gauche, bas, etc.). Notons qu'avec ce type de gestes, on retrouve la logique des *Marking Menus* [Kurtenbach et Buxton, 1991]. La manière de tracer les dix chiffres et les huit directions est illustrée dans la Figure 6.8.

Outre la nature, directionnelle ou numérale du matériel, nous avons manipulé dans l'expérience le mode de présentation du matériel (voir la Figure 6.7). Le matériel, chiffre ou simple direction, pouvait être présenté soit sous forme graphique comme dans la Figure 6.8, auquel cas il s'agissait de recopier la forme, direction ou chiffre, affichée à l'écran, soit sous forme textuelle, auquel cas le tracé devait être synthétisé ex nihilo par le participant.

Il est important de remarquer que dans cette étude nous ne nous intéressons pas à la qualité des tracés mais à la direction spontanée des tracés.

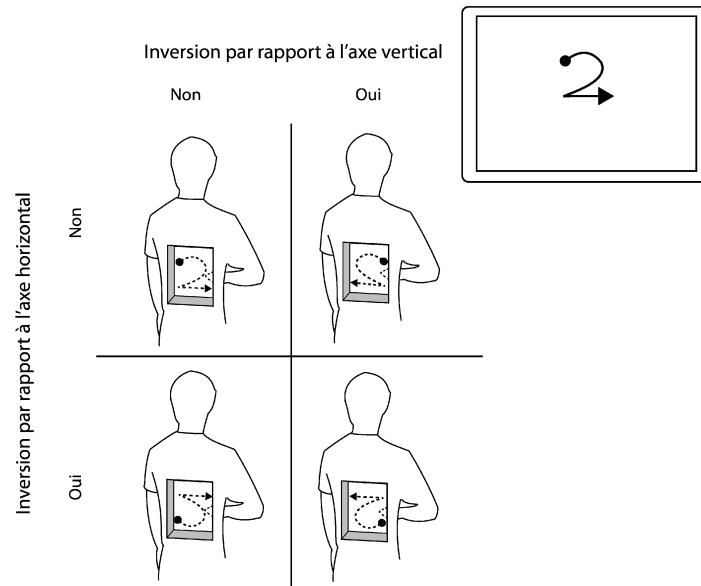


FIGURE 6.6: Représentations internes de l'orientation spatiale identifiées chez les participants : exemple du chiffre "2"

		Matériel présenté	
		Direction	Chiffre
Mode de Présentation	Graphique		
	Textuel	<b>Bas Droite</b>	<b>Deux</b>

FIGURE 6.7: Exemple de stimuli utilisés pour le mode de présentation et le matériel présenté. Le point indique le point de départ.

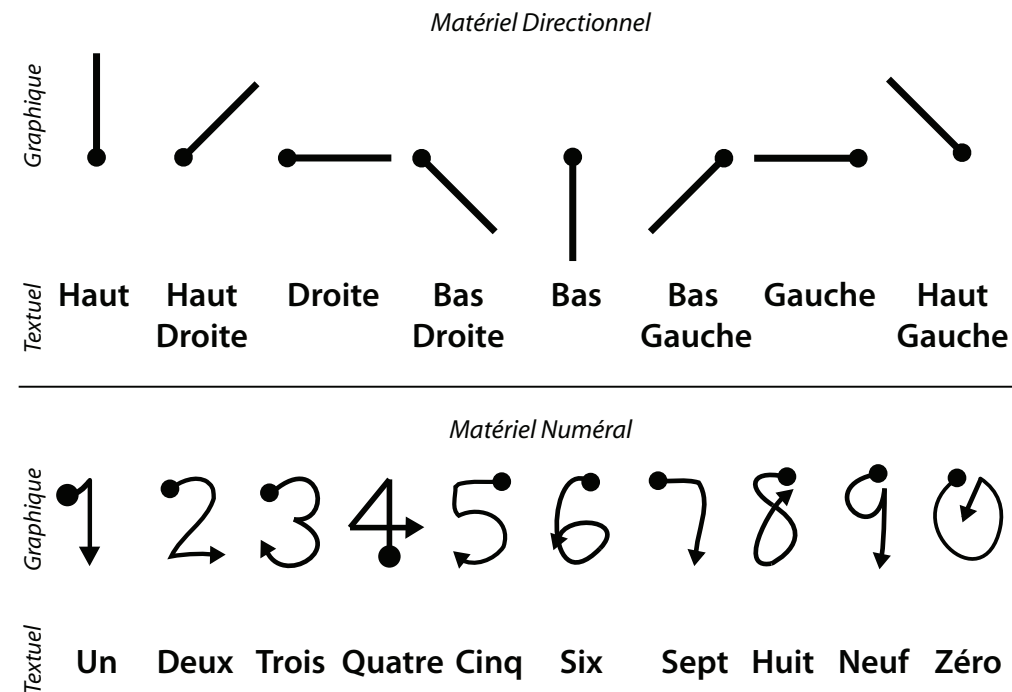


FIGURE 6.8: Stimuli utilisés pour l'expérimentation : dans la partie supérieure, les stimuli pour le matériel directionnel ; dans la partie inférieure, les stimuli pour le matériel numéral

### 6.2.6 Équipement et procédure

Nous avons utilisé Microsoft Kinect for Xbox 360<sup>1</sup> pour capturer les gestes de la main des participants sur leur abdomen. Bien que les participants se tenaient debout en face de l'écran à une distance optimale pour le bon fonctionnement de Kinect, la limite de la technologie ne permet pas de capturer pleinement les légères différences de profondeur entre la main et l'abdomen lorsqu'ils sont en contact. Pour améliorer le suivi du mouvement, nous avons habillés les participants de t-shirts blancs. Le logiciel permettant de contrôler l'expérimentation a été exécuté sur un MacBook Pro doté d'un processeur Intel Core 2 Duo cadencé à 2.66 GHz. Ce logiciel a été implémenté avec le support des bibliothèques Qt<sup>2</sup>. L'écran sur pied utilisé pour afficher les stimulus et les informations de l'expérimentation était un écran de 64 pouces avec une résolution de 1024×768 pixels. Une Magic Mouse Apple contrôlait l'état d'enregistrement du geste.

Le participant était positionné debout à environ 3 mètres de l'écran. L'expérimentateur était assis derrière le participant, à une distance de 2 mètres environ. Avant le début de l'expérimentation, les participants ont été instruits de réaliser un geste continu dès le contact de la main avec l'abdomen pour chaque essai. Pour la présentation graphique du stimulus, il fallait le reproduire exactement le symbole visible sur l'écran. Pour le mode de présentation textuelle, nous avons demandé aux participants de réaliser le geste qui représenterait le mieux l'image de leur représentation mentale associée au stimulus.

Pendant l'expérimentation, les participants n'ont pas été autorisés à observer leur abdomen car nous souhaitons que les gestes des participants ne reposent que sur des retours proprioceptifs. Afin de leur laisser un temps de réflexion suffisant pour correctement définir leur représentation de l'orientation spatiale, ils disposaient d'autant de temps que nécessaire pour réaliser leur geste. Les stimulus ont été présentés au centre de l'écran. Aucun retour visuel n'a été affiché sur l'écran pendant la réalisation du geste.

Pour délimiter les gestes, nous avons équipé les participants d'une souris sans fil pour signaler le contact de la main avec la surface du ventre, Kinect ne permettant pas de détecter de manière fiable ce contact. Pour effectuer le geste, les participants devraient d'abord poser leur main dominante sur leur abdomen, puis quand ils étaient prêts, appuyer sur le bouton de la souris

---

1. <http://www.xbox.com/fr-CA/Kinect>

2. <http://qt-project.org/>

tenue par la main non dominante, et réaliser le geste. Le geste était considéré comme achevé lorsque les participants relâchaient le bouton de la souris pour arrêter l'enregistrement des données. Le relâchement du bouton de la souris déclenchait également l'essai suivant. Les participants pouvaient s'entraîner pour chaque geste pendant une série avant l'expérimentation. Ils ont aussi été autorisés à faire une pause après chaque série s'ils le désiraient. Le tracé de chaque geste a été enregistré par notre système et nous avons enregistré chaque geste avec une caméra vidéo.

L'ordre des deux facteurs a été contrebalancé en utilisant un Carré Latin partiel. Pour chaque condition, les participants ont réalisé les 18 gestes différents pour chacune des cinq séries. Pour résumer, le plan expérimental correspondait à :  $12 \text{ participants} \times 2 \text{ conditions} \times 5 \text{ séries} \times 18 \text{ gestes}$  (8 directions + 10 chiffres) = 2160 échantillons. A la fin de l'expérimentation, un questionnaire a été distribué pour récolter des informations sur la stratégie suivie pour effectuer les gestes.

### 6.2.7 Résultats

Avant de procéder à une analyse approfondie de nos données, nous avons écarté 8,27% de nos échantillons où l'orientation n'était pas identifiable à cause des propriétés symétriques de certains gestes (pour les chiffres 0, 1 et 8) ou à cause des erreurs du dispositif d'acquisition. Nous avons classé les échantillons manuellement selon quatre classes différentes de la représentation spatiale mentale de l'orientation (RSMO). Ces classes correspondent aux classes d'inversion symétriques horizontales et verticales discutées auparavant (Figure 6.6).

Pour investiguer la RSMO, nous considérons deux types d'information. D'abord, nous nous intéressons à la stabilité mentale de la représentation qui donne une information sur la consistance du choix de la stratégie des utilisateurs pour interagir avec leur abdomen. Puis, nous étudions l'orientation de ces représentations pour comprendre comment elle pourrait influencer la réalisation des gestes.

#### 6.2.7.1 Stabilité mentale de la représentation et entropie

Dans le domaine de la théorie de l'information, l'entropie est une grandeur qui mesure la quantité d'incertitude qu'un événement aléatoire se produise. Cette quantité est équivalente à la quantité d'information contenue dans un

message et se mesure en bits [Shannon, 1948]. Dans notre situation, nous nous en servons pour mesurer la quantité d'incertitude du choix des utilisateurs d'une RMSO parmi les quatre identifiées précédemment (Figure 6.6). L'entropie  $H$  d'une variable discrète  $x$  (par exemple le choix d'une RSMO), d'un ensemble de  $N$  événements parmi  $\{x_1 \dots x_n\}$  (parmi les quatre représentations) est exprimée par la formule de Shannon [Shannon, 1948] :

$$H(x) = \log_b(N) - \frac{1}{N} \sum_i n_i \cdot \log_b(n_i) \quad (6.1)$$

La valeur de l'entropie  $H$  peut être positive ou nulle. Si cette valeur est nulle, alors la probabilité que l'événement se réalise est maximale. Dans notre cas, cela signifierait que parmi les quatre choix possibles de représentation, les utilisateurs choisiraient toujours la même. Lorsque les choix de la représentation sont équiprobables, alors  $H$  atteint sa valeur maximale et traduit une situation avec une incertitude maximale. Ce cas illustrerait le fait que les utilisateurs choisissent l'une des quatre représentations de manière imprévisible. Avec quatre représentations différentes, la valeur de l'entropie s'exprime sur deux bits d'information avec une valeur comprise entre 0 et 2.

Dans le mode de présentation graphique, la valeur moyenne de l'entropie était légèrement différente pour le cas des directions ( $\bar{H} = 0,27 \text{ bit}$ ,  $\sigma = 0,34 \text{ bit}$ ) et celui des chiffres ( $\bar{H} = 0,32 \text{ bit}$ ,  $\sigma = 0,40 \text{ bit}$ ). Dans le mode de présentation textuel, la différence du cas avec les directions ( $\bar{H} = 0,33 \text{ bit}$ ,  $\sigma = 0,40 \text{ bit}$ ) et celui des chiffres ( $\bar{H} = 0,65 \text{ bit}$ ,  $\sigma = 0,45 \text{ bit}$ ) est plus importante (Figure 6.9). Aucun effet significatif du type de matériel, du mode de présentation, ou de l'interaction entre les deux facteurs n'a été relevé.

Ces premiers résultats, bien que non significatifs révèlent une tendance. Dans le mode de présentation graphique, la différence du degré d'incertitude moyen entre les gestes directionnels et numériques est très faible ( $\Delta = 0,05 \text{ bit}$ ), contrairement au mode de présentation textuelle où elle est deux fois supérieure ( $\Delta = 0,32 \text{ bit}$ ). Si maintenant nous considérons principalement les types de matériel, nous observons une très faible différence de la moyenne de l'incertitude pour le matériel directionnel entre les deux modes de présentation ( $\Delta = 0,06 \text{ bit}$ ). En revanche, on remarque une différence importante pour le matériel numérique ( $\Delta = 0,33 \text{ bit}$ ).

A la vue de ces résultats préliminaires, la quantité élevée d'incertitude dénote une représentation mentale plus instable pour les chiffres et, particulièrement dans le mode de présentation textuelle. Enfin, il est intéressant de noter la quasi-stabilité de la représentation mentale pour les gestes directionnels



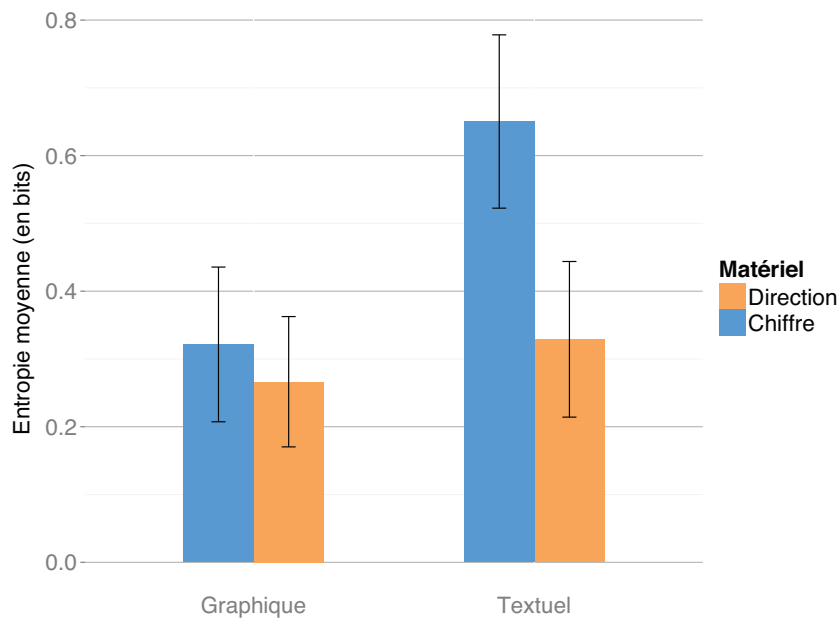


FIGURE 6.9: Entropie moyenne (en bits) en fonction du mode de présentation et du matériel présenté. Les barres d'erreurs représentent un intervalle de confiance à 95%.

dans les deux modes de présentation. Ces observations prometteuses nous poussent à approfondir notre investigation et étudier précisément ces représentations mentales.

#### 6.2.7.2 Représentation spatiale mentale de l'orientation

Les résultats précédents montrent une instabilité variable de la représentation mentale selon les conditions de l'expérimentation. Afin de caractériser cette instabilité, nous proposons d'examiner ces représentations en terme d'inversion par rapport à l'axe horizontal et vertical relative à la représentation de référence illustrée à la Figure 6.5 comme expliqué au début de la section 6.2.5. La Figure 6.6 présente les différentes représentations qui vont être discutées et leur relation spatiale par rapport à la représentation de référence. Le cadran nord-ouest représente la représentation de référence. Le cadran nord-est représente une inversion par rapport à l'axe vertical alors que le cadran sud-ouest de la Figure représente une inversion par rapport à l'axe horizontal. Enfin, le dernier cadran représente une inversion par rapport, à la fois, à l'axe

horizontal et vertical de la représentation.

Pour chaque condition expérimentale, les fréquences de la distribution de chaque transformation des échantillons ont été calculées, ainsi que la moyenne entre les participants.

En moyenne, 13,5% des échantillons étaient une inversion par rapport à l'axe vertical toutes conditions confondues. Spécifiquement, concernant le matériel numéral, une inversion de ce type a été relevée pour 6,4% des échantillons pour la condition de présentation graphique et 42,9% pour l'autre condition. Pour le matériel directionnel, 1,8% des échantillons présentaient ce type d'inversion lorsque le stimulus était présenté graphiquement et 2,9% dans l'autre mode de présentation. Une ANOVA a révélé un effet significatif du type de matériel ( $F_{1,11} = 15,87$ ;  $p < 0,0022$ ), un effet significatif du mode de présentation ( $F_{1,11} = 11,27$ ;  $p < 0,0065$ ), et un effet significatif de l'interaction de ces deux facteurs ( $F_{1,11} = 9,78$ ;  $p < 0,0097$ ). Les résultats sont illustrés dans la Figure 6.10.

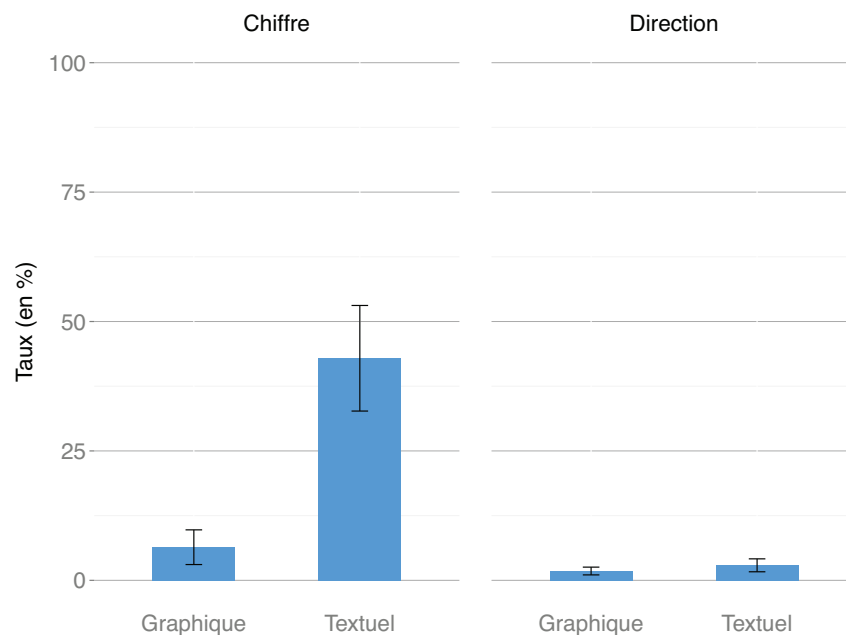


FIGURE 6.10: Taux des cas d'inversion par rapport à l'axe vertical. Les barres d'erreur représentent l'intervalle de confiance à 95%.

Transformation		Vertical				Horizontal			
Participants	Matériel	Direction		Chiffre		Direction		Chiffre	
	Présentation	Textuel	Graphique	Textuel	Graphique	Textuel	Graphique	Textuel	Graphique
	$P_1$	0,079	0,079	0,923	0,000	0,000	0,000	0,000	0,000
	$P_2$	0,000	0,000	0,265	0,182	0,000	0,000	0,000	0,000
	$P_3$	0,000	0,027	0,531	0,400	0,000	0,054	0,000	0,033
	$P_4$	0,025	0,026	0,844	0,029	0,000	0,000	0,000	0,000
	$P_5$	0,128	0,000	0,429	0,000	0,026	0,000	0,029	0,000
	$P_6$	0,028	0,056	0,000	0,029	0,806	0,778	1,000	0,971
	$P_7$	0,087	0,030	0,222	0,097	0,130	0,030	0,000	0,000
	$P_8$	0,030	0,000	0,000	0,000	0,758	0,750	1,000	1,000
	$P_9$	0,000	0,000	0,606	0,000	0,000	0,000	0,000	0,000
	$P_{10}$	0,000	0,000	0,037	0,000	0,000	0,000	0,852	0,000
	$P_{11}$	0,000	0,000	0,200	0,029	0,000	0,000	0,143	0,235
	$P_{12}$	0,000	0,000	1,000	0,037	0,000	0,000	0,000	0,000

TABLE 6.1: Fréquence moyenne des transformations pour chaque participant et pour chaque condition expérimentale

En ce qui concerne les inversions par rapport à l'axe horizontal, nous avons relevé 17,5% des échantillons en moyenne pour toutes conditions confondues. Plus précisément, pour le matériel numéral, 18,9% présentaient également ce type d'inversion pour la présentation graphique et 24,5% pour l'autre mode de présentation. Enfin, 13,5% des échantillons des gestes directionnels présentaient également de telles inversions dans le cas d'une présentation graphique et 13,1% si elle était textuelle. Les résultats sont illustrés dans la Figure 6.11.

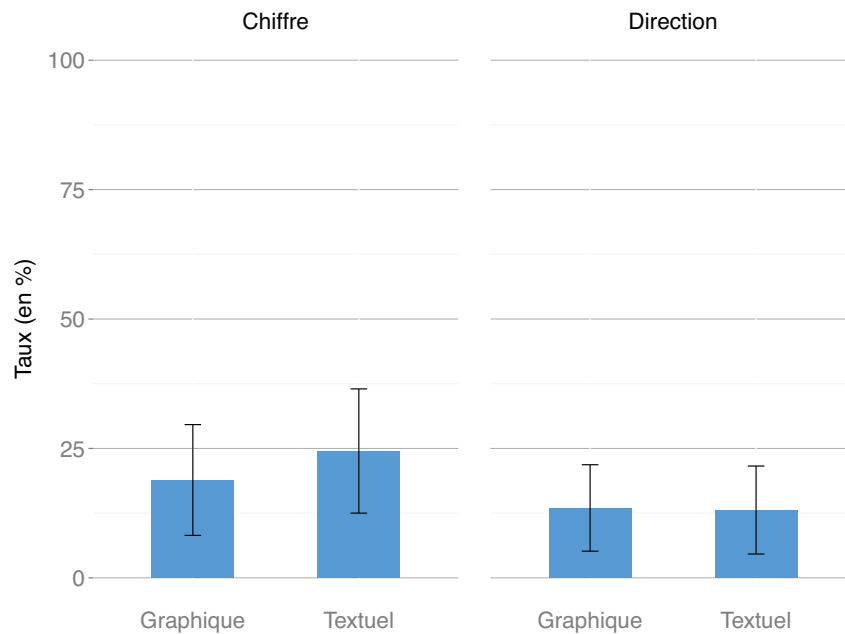


FIGURE 6.11: Taux des cas d'inversion par rapport à l'axe horizontal. Les barres d'erreur représentent l'intervalle de confiance à 95%.

Afin d'obtenir un indicateur de la charge cognitive éprouvée par les participants, nous avons mesuré le temps de réaction. Ce temps correspond à la durée entre l'affichage du stimulus et la pression du bouton de la souris permettant d'activer l'enregistrement du geste. Dans la condition du mode de présentation graphique, le temps moyen de réaction atteint 1489 millisecondes (ms) pour le matériel de type directionnel et 2277 ms pour celui de type numéral. Dans la condition de présentation textuelle, le temps moyen de réaction s'élève à 1548 ms pour le matériel de type directionnel et 1949 ms pour celui de type numéral. Une ANOVA montre un effet significatif du facteur type de matériel ( $F_{1,11} = 8,22$ ;  $p < 0,007$ ) tandis qu'aucun effet si-

gnificatif du facteur de mode présentation et de leur interaction n'a été relevé.

Parce que nous avons instruit les participants de prendre le temps nécessaire pour effectuer leur geste, le temps d'exécution du geste n'est pas représentatif de la performance réelle de l'accomplissement du geste.

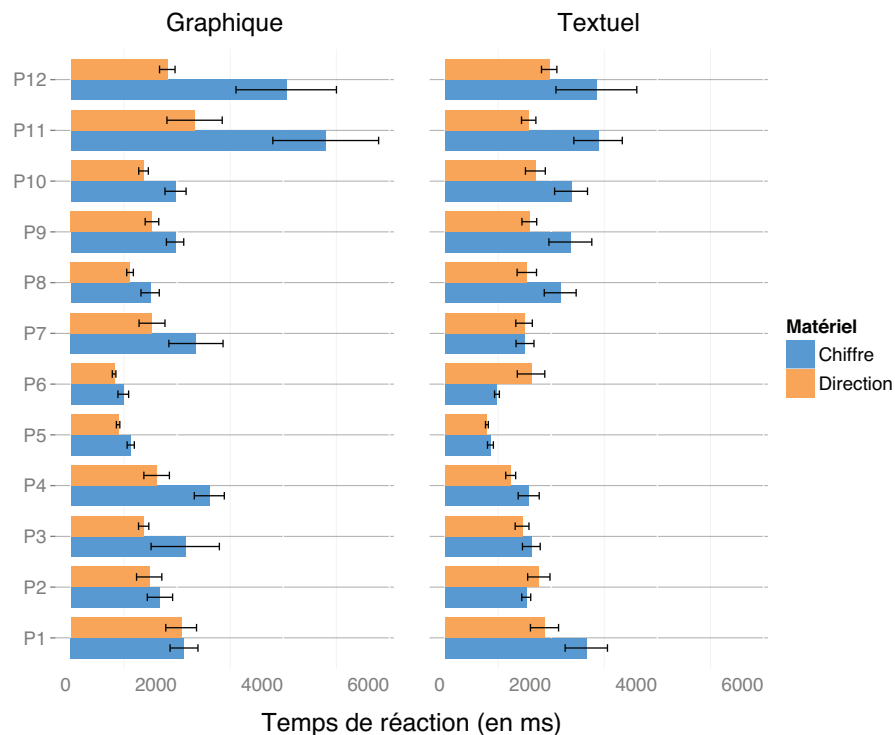


FIGURE 6.12: Temps moyen de réaction par participant. Les barres d'erreur représentent l'intervalle de confiance à 95%.

### 6.2.7.3 Discussions sur les inversions

En général la propension à l'inversion par rapport à l'axe horizontal est plus importante que celle de l'inversion par rapport à l'axe vertical, excepté quand le matériel numéral est présenté de façon textuelle

Dans le cas de l'inversion par rapport à l'axe horizontal, les taux se rangent entre 20 et 25% quelque soit la condition expérimentale. Ce résultat à la fois stable et surprenant est intéressant tant la littérature psychologique suggère que la représentation spatiale de la direction vers le haut et vers le bas est guidée par la perception de la gravité par le système vestibulaire, le système

nerveux somatique et la vision [Mergner et Rosemeier, 1998]. Cette inversion de la direction haut avec celle du bas peut être expliquée par la représentation mentale de l'utilisateur de la configuration spatiale de son corps et de sa tête par rapport au sol [Mergner et Rosemeier, 1998]. Certains utilisateurs ont mentionné réaliser le geste en s'imaginant regarder l'espace d'interaction : « Je n'ai pas réfléchi au sens de l'écriture. J'ai agi naturellement. Il est fort probable que j'écrive sur mon ventre comme si je regardais mon ventre », ou encore « [...] comme si je regardais mon ventre. Parce que ça m'a semblé le plus facile si il n'y a pas d'opération géométrique miroir à effectuer ».

Le cas de l'inversion par rapport à l'axe vertical est également intéressant. D'une manière générale, le taux d'inversion reste significativement très faible dans le cas des directions. Cette tendance est aussi remarquable pour les chiffres, mais seulement lorsqu'ils sont présentés graphiquement. Au contraire, le mode de présentation textuelle montre un taux d'inversion plus de 6,5 fois supérieur au mode de présentation graphique.

Plusieurs raisons peuvent expliquer cette différence importante. Premièrement, la présentation graphique des stimuli donne un référentiel sur lequel les participants peuvent s'appuyer pour réaliser leur geste. En effet dans cette condition, un stimulus est composé d'un point de départ, d'un point d'arrivée et d'un sens de parcours. Ces indications soulagent l'effort cognitif nécessaire à planifier le parcours du geste. Présenté textuellement, le stimulus ne fournit aucune information pour sa reproduction obligeant les participants à se représenter mentalement le symbole avant de le dessiner. Par exemple un participant explique : « Oui j'ai réfléchi au sens de l'écriture car je n'ai pas l'habitude d'écrire dans ce sens. Par contre, je n'ai pas réfléchi lorsque le chiffre était dessiné à l'écran ».

Trois de nos participants ont changé de représentation pendant l'expérience pour la condition matériel numéral et mode de présentation textuel. Certains ont soulevé la difficulté de choisir une représentation unique : « Le sens est venu naturellement, mais il a pu varier en cours d'expérience sans que j'y prête attention ». Un autre participant a affirmé : « J'ai essayé de faire attention mais pour le 3 en particulier je n'ai pas réussi à me décider pour un sens. Pour 5 et 2, ça allait mieux ». Enfin, la durée de réaction traduit également la demande cognitive importante pour le matériel numéral avec environ 400 ms supplémentaires en moyenne pour les chiffres quand ils sont présentés textuellement.

Les données qualitatives suggèrent que le choix de la représentation interne de l'orientation spatiale est liée à la représentation mentale de la situation des

participants. A la question « *Pendant que vous étiez en train de réaliser les gestes, avez-vous réfléchi au sens de l'écriture ? Si oui, quel est le sens que vous avez choisi et pourquoi ?* », le participant 1 répond : « *Dans le cas des stimulus en lettres, j'ai réalisé le geste comme si j'étais face au miroir sans trop réfléchir. Dans le cas des tracés, j'ai reproduit le tracé en miroir par rapport à l'écran* ». Le participant 2 a répondu : « *J'ai essayé de faire comme si je traçais les lettres face à un miroir* ». Participant 4 dit : « *Je me suis imaginé à la place d'un spectateur (en face de moi)* ». Le participant 12 affirme : « *For me it was easier manner to perform (draw) from the device point of view* ».

L'hétérogénéité des réponses souligne des difficultés à choisir une représentation interne spatiale unique lorsque les gestes à réaliser sans regarder sont relativement complexes. Cependant, si le concepteur décide de choisir d'imposer une représentation mentale aux utilisateurs, ces images mentales peuvent guider les utilisateurs pour choisir ou pour indiquer la perspective dans laquelle l'utilisateur doit se projeter pour réaliser les gestes.

Contrairement au matériel numéral, les résultats pour les gestes directionnels sont relativement stables. La proportion des inversions verticales est quasi-nulle quel que soit le mode de présentation. Étonnamment, bien qu'aucune règle physique, telle que la gravité pour la représentation directionnelle haut / bas par exemple, ne semble gouverner la représentation gauche / droite, cette stabilité augure une représentation mentale commune au sein des participants.

Étant donné les résultats obtenus pour le matériel de type numéral dans le cas d'une présentation textuelle, la disparité des inversions ne permet pas de généraliser le comportement des utilisateurs pour une prédiction efficace du choix de l'orientation des gestes. Cependant, nous avons identifié des images mentales qui pourraient permettre de guider les utilisateurs lors d'un choix d'une orientation spatiale pour la réalisation des gestes sur l'abdomen. Cette piste mérite des investigations supplémentaires. Une solution alternative consisterait d'intégrer le geste et ses représentations selon chaque type d'inversion dans le vocabulaire à apprendre par le système de reconnaissance. Il est important de noter la durée relativement importante des temps de réaction pour le matériel numéral et le mode de présentation textuel. Cette charge cognitive pourrait frustrer l'utilisateur particulièrement dans un environnement où son objectif est de se reposer et de se divertir.

Les résultats obtenus pour le matériel de type directionnel présentent une stabilité remarquable en terme d'inversion gauche / droite. Bien que moins



tranchants, les résultats concernant l'inversion horizontale présente également une stabilité qui reste exploitable. Par ailleurs, les temps de réaction liés au matériel de type directionnel montrent une durée moyenne nettement en deçà des durées observées pour le matériel de type numéral. Cette différence suppose un effort cognitif inférieur quant à la construction de la représentation spatiale mentale de l'orientation. Ce bilan prometteur comparé aux résultats observés pour la condition matériel de type numéral nous incite à approfondir notre investigation. La partie suivante est consacrée à l'étude des gestes effectués par les participants pour la condition du matériel de type directionnel.

#### 6.2.8 *Analyse de l'exécution des gestes directionnels sur l'abdomen*

Nous ignorons dans cette section le matériel numéral pour se focaliser sur le matériel directionnel. Le geste auquel nous nous intéressons ici consiste dans le simple tracé d'un segment de droite sur la surface de l'abdomen. Rappelons que l'expérience faisait intervenir deux variables indépendantes, le mode de présentation du matériel (graphique vs. textuel) et la direction demandée (huit directions, dont quatre canoniques et quatre diagonales). L'analyse présentée ici a été réalisée sur les échantillons ne présentant aucune inversion dans l'analyse des représentations mentales. Ce cas représente le groupe d'échantillon le plus important parmi les quatre discutés dans l'analyse précédente.

Nous avons considéré quatre variables dépendantes de base, définies au niveau du mouvement individuel :

- Le *temps d'exécution du geste*, égal à la durée de la pression exercée avec la main non-dominante sur le bouton de souris sans fil.
- L'*amplitude du geste*, calculée comme la distance linéaire séparant les positions de la main aux moments de la pression et du relâchement du bouton de souris.
- La *linéarité* du tracé, estimée par le coefficient de détermination  $r^2$  de la régression linéaire effectuée sur les coordonnées en x et en y des positions successives de la main pendant le geste. La valeur de cet indice n'étant naturellement pas indépendante de l'orientation du tracé dans le plan, en pratique nous avons estimé la linéarité comme la valeur maximale atteinte par le  $r^2$  lorsque le tracé était pivoté sur  $360^\circ$ .

- L'*erreur directionnelle* définie comme la différence entre l'orientation de la ligne passant par le point de départ et par le point d'arrivée du mouvement et la direction demandée par le stimulus.

À partir de la mesure de l'erreur directionnelle, nous avons estimé deux statistiques :

- Le *biais directionnel* (ou erreur systématique) du tracé, qui correspond à l'erreur directionnelle moyenne du geste
- La *variabilité directionnelle* (ou erreur variable) du tracé, calculée comme la moyenne des valeur absolue de l'erreur directionnelle.

L'objectif de cette expérience exploratoire était de déterminer comment les caractéristiques des gestes varient en fonction de leur direction, pour chacun des deux modes de présentation. Les données que nous avons recueillies permettent une première évaluation de l'utilisabilité de ces différents gestes dans un vocabulaire de commandes d'entrée employant l'abdomen comme surface interactive.

#### 6.2.8.1 Résultats

Pour présenter les données de cette expérience, nous adopterons la convention de représentation définie plus haut dans la Figure 6.5, où le plan de l'écran est mis en correspondance directe, sans inversion, avec celui de l'abdomen du participant, vu depuis l'arrière : ainsi par exemple, nous dirons que le geste est orienté vers l'Est s'il est orienté vers la droite à la fois dans le plan de l'écran et dans celui de l'abdomen du participant.

La variable représentée dans la Figure 6.13 est le temps d'exécution calculé sur l'ensemble des participants, pour chacun des deux mode de présentation et pour chacune des huit directions dans laquelle le geste devait être effectué. Nous n'observons pas de différences remarquables entre les directions, les gestes étant dans tous les cas exécutés en un temps de l'ordre de 1,2 seconde. Néanmoins les temps tendent à être légèrement plus courts pour les directions canoniques (du moins du côté de l'Est, du Nord et de l'Ouest) que pour les directions diagonales. Le mode de présentation du matériel directionnel n'exerce pas d'influence notable sur cette variable dépendante.

L'amplitude varie peu selon la direction du geste, le mouvement s'étendant sur une distance comprise entre 15 et 20 cm (Figure 6.14). Il n'y a pas d'effet

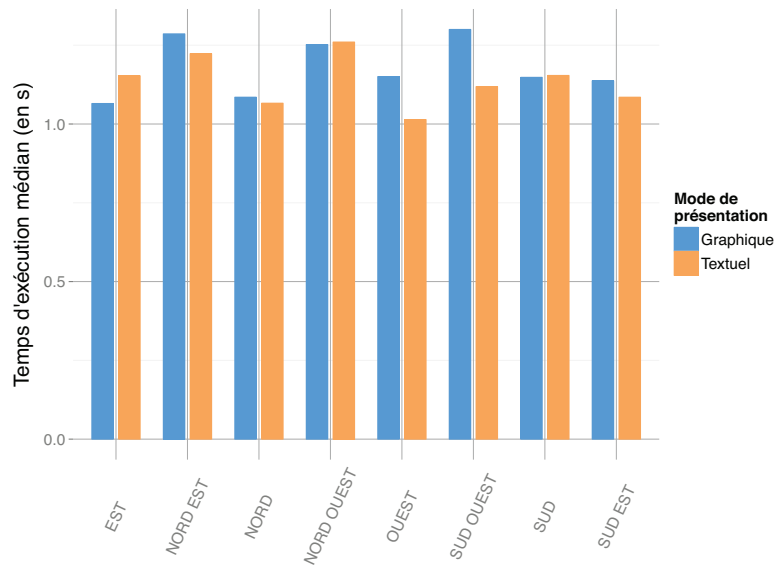


FIGURE 6.13: Temps d'exécution médians en secondes.

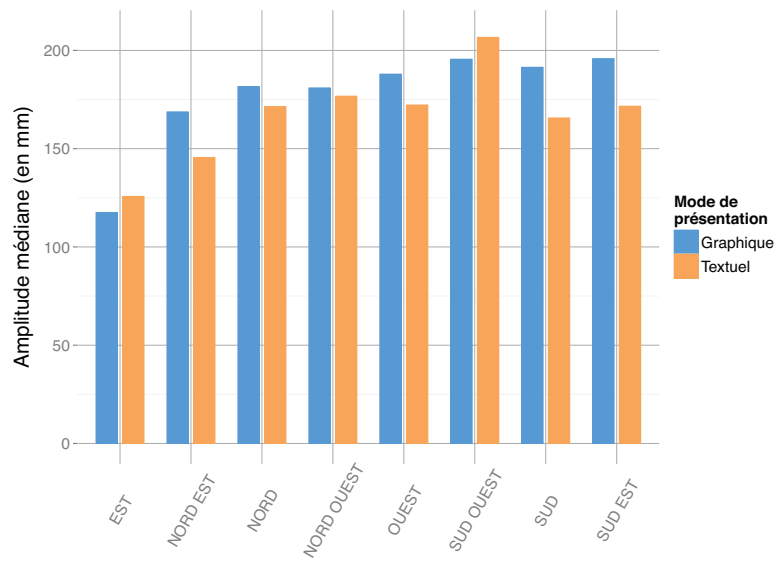


FIGURE 6.14: Amplitude médiane du geste en millimètres.

notable du mode de présentation du matériel.

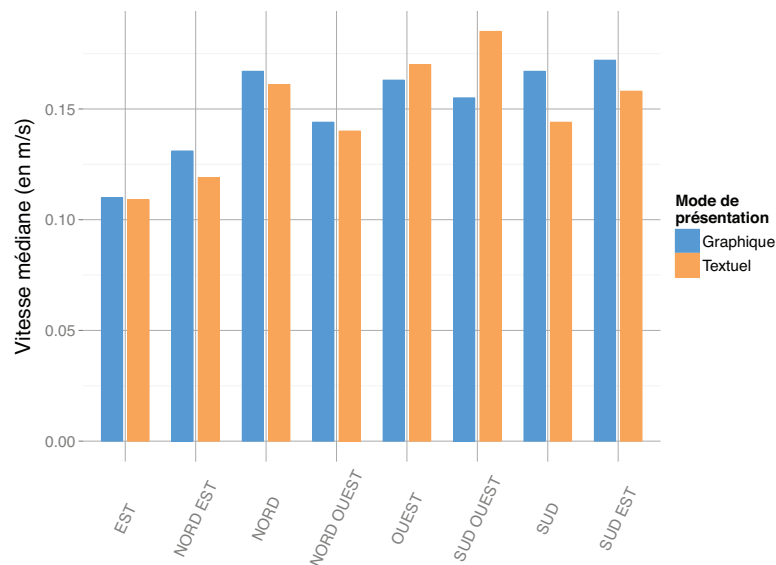


FIGURE 6.15: Vitesse médiane du geste (en mètres par seconde).

En combinant l'information d'amplitude et l'information de temps contenues dans les Figures 6.13 et 6.14, on peut calculer la vitesse moyenne du mouvement (Figure 6.15). La médiane de cette vitesse, de l'ordre de 15 centimètres par seconde, apparaît relativement constante dans les huit directions et pour les deux modes de présentation. Le mode de présentation du matériel est une fois encore sans effet. Comme cela est le cas dans chacune des figures qui suivent, nous prenons ci-dessous le parti de mélanger les données correspondant aux deux modes de présentation.

La Figure 6.16 montre la linéarité moyenne du tracé en fonction de la direction du geste. On constate que les tracés orientés vers l'Ouest ( $135^\circ$ ,  $180^\circ$  ou  $225^\circ$ ), qui reposent sur des gestes en poussée du bras avec extension du coude, sont plus linéaires que ceux orientés vers l'Est ( $315^\circ$ ,  $0^\circ$  ou  $45^\circ$ ), qui demandent une flexion du coude. Cet effet, présent chez neuf participants sur dix, est nettement significatif statistiquement selon un simple test du signe ( $p < 0.05$ ).

L'examen de l'erreur systématique (Figure 6.17) suggère que les directions Est, Nord-Est et Nord sont plus difficiles à exécuter que les autres. Les gestes vers l'Est tendent à être biaisés vers le Sud-Est (biais négatif), tandis que les gestes vers le Nord-Est et le Nord tendent à être biaisés dans le sens antihor-

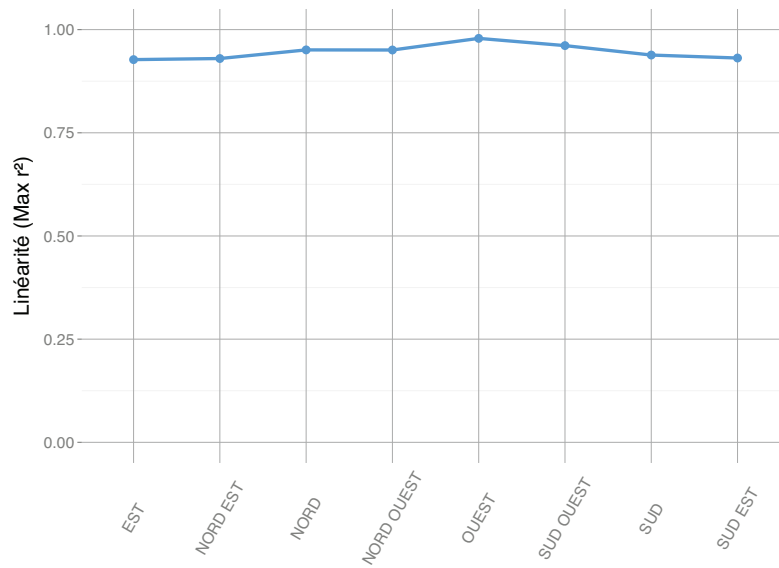


FIGURE 6.16: Linéarité moyenne du geste : valeur moyenne des maxima du coefficient de détermination  $r^2$  de la régression linéaire des tracés pivotés sur  $360^\circ$ .

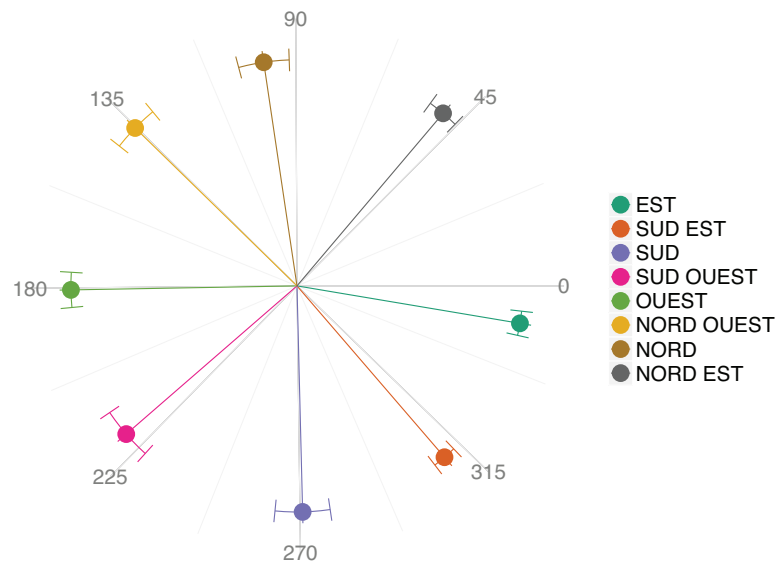


FIGURE 6.17: Erreur directionnelle moyenne du geste (en degrés). Les barres d'erreur représentent l'intervalle de confiance à 95%.

raire (biais positifs).

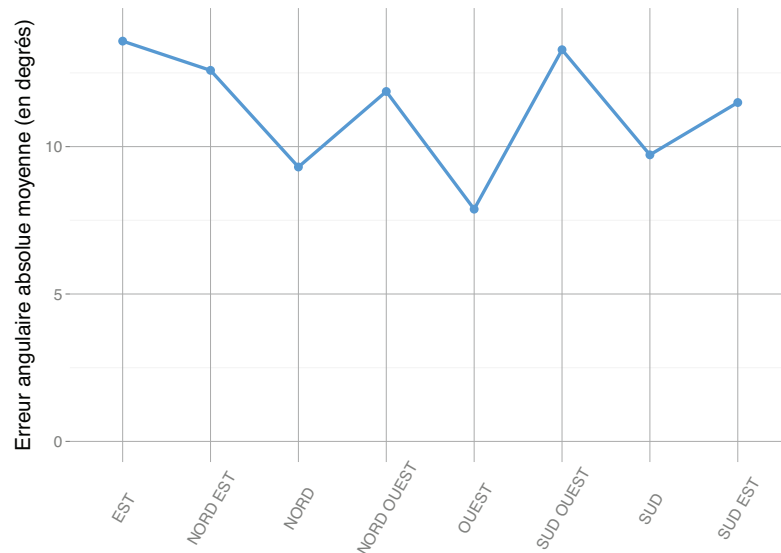


FIGURE 6.18: Variabilité directionnelle du geste (en degrés) : moyenne des valeurs absolues de l'erreur directionnelle

La variabilité angulaire, ou erreur variable directionnelle du tracé (Figure 6.18) montre une oscillation qui reflète une différence entre les directions canoniques et les directions diagonales. Avec une exception pour la direction Est, les participants ont été plus constants dans la reproduction de leurs gestes quand on leur a demandé de les tracer dans une direction canonique (où la moyenne des erreurs angulaires absolues est de  $10.1^\circ$ ) plutôt que dans une direction diagonale ( $12.3^\circ$ , soit un accroissement de 20%). La différence étant de même signe sur 9 des 10 participants, l'effet est nettement significatif ( $p < 0.05$ ).

#### 6.2.8.2 Discussion

Cette recherche avait pour objectif d'évaluer, selon une série de critères complémentaires, l'utilisabilité de simples gestes de tracés linéaires dans huit directions sur l'abdomen. Les résultats indiquent que l'espace d'interaction considéré n'est pas entièrement homogène.

Avec des  $r^2$  moyens compris entre 0.93 et 0.98, nous avons vu que les tracés produits par les participants étaient à peu près linéaires. Toutefois, le fait que ces tracés paraissent un peu plus linéaires vers l'Ouest que vers l'Est donne à penser que les directions OUEST devraient être utilisées en priorité dans des implémentations de ce genre de technique lorsque l'utilisateur agit avec la main droite. Il serait intéressant de comparer ces résultats avec ceux d'une étude basée sur des gestes de la main gauche.

Les résultats de notre étude montrent que les erreurs systématiques, dont la valeur est généralement très inférieure à  $10^\circ$ , sont suffisamment minimales pour ne pas empêcher la distinction d'une direction parmi huit. Les gestes ne sont généralement biaisés que de quelques degrés, à l'exception de ceux orientés dans une région angulaire allant de  $0^\circ$  à  $90^\circ$  environ. Ces observations confirment que la technique est potentiellement utilisable.

Une variabilité angulaire de l'ordre de  $10^\circ$  a été observée parmi les tracés des gestes des participants. Les résultats indiquent également que la reproductibilité des gestes orientés semble meilleure dans les directions canoniques. Ces résultats sont en mesure de guider les choix des concepteurs : par exemple, il sera préférable de privilégier une disposition canonique dans le cas d'un vocabulaire directionnel à quatre choix.

#### 6.2.9 Conclusion

Dans ce chapitre, nous nous sommes intéressés aux gestes réalisés sur la surface du corps pour interagir avec la télévision. Ces gestes sont par exemple utiles lorsque la télécommande est hors d'usage ou difficilement accessible. Plus particulièrement, nous avons proposé d'utiliser la surface de l'abdomen comme surface d'entrée pour l'interaction. Les caractéristiques particulières de cette surface, comme sa taille, sa stabilité ou encore son exposition limitée aux contacts accidentels, lui confèrent de nombreux avantages pour l'interaction gestuelle. Cependant, l'orientation et la position de l'abdomen par rapport à la tête impliquent pour les utilisateurs de se passer de contrôle visuel et de se fier uniquement aux retours proprioceptifs offerts par le mouvement des autres parties du corps et leurs contacts avec l'abdomen. Pour étudier la capacité à réaliser ces gestes, nous abordons le problème selon deux approches : la préparation du geste et son exécution.

Suite à l'enregistrement des tracés des gestes directionnels et de type numérique réalisés sur la surface du ventre, nous avons d'abord classifié les échan-

tillons selon leur orientation. Parmi ces échantillons, nous avons identifié quatre représentations mentales spatiales de l'orientation différentes.

La première partie a été consacrée, d'une part, à une revue des travaux s'intéressant à l'interaction sur le corps et, d'autre part, la description des propriétés de l'abdomen à considérer si l'on envisage son utilisation comme une surface interactive.

La seconde partie de ce chapitre s'est concentrée sur l'aspect cognitif lié à la capacité des utilisateurs à réaliser les gestes. La première étude se consacre à l'observation de la stabilité de la représentation spatiale mentale de l'orientation des utilisateurs lorsqu'ils réalisent des gestes sur la surface de l'abdomen. Les résultats montrent que cette représentation est stable lorsque les participants sont guidés par une représentation graphique définie par un référentiel spatial. Dans le cas où aucune information spatiale ne leur est fournie, bien que la représentation est instable pour les gestes complexes (chiffres), elle reste stable pour des gestes simples (directionnels).

Une seconde analyse plus approfondie de ces représentations montrent que la réalisation des gestes complexes (basés sur des chiffres) sont sujets à une probabilité d'inversion importante. Ces inversions sont liées à la représentation dans laquelle les utilisateurs se projettent avant de réaliser le geste. Ces images mentales pourraient guider les utilisateurs si le concepteur choisit d'imposer une représentation spatiale de l'orientation particulière. Par ailleurs, contrairement aux gestes simples (basés sur la direction), ces gestes demandent un effort cognitif important. Le cas des gestes simples semble plus prometteur et pour cette raison il fait l'objet de la dernière partie de ce chapitre.

Enfin, la dernière partie de ce chapitre s'est focalisée sur l'aspect physique de la réalisation du geste. Plus particulièrement, l'objet de cette partie est l'analyse de l'exécution des gestes directionnels sur la surface de l'abdomen. L'analyse des tracés selon des critères élémentaires du mouvement montrent qu'une technique d'interaction basée sur ces gestes serait potentiellement utilisable. Les erreurs systématiques angulaire sont assez petites pour pouvoir facilement discrétiser huit directions. Lorsque les utilisateurs interagissent avec la main droite, nos résultats révèlent que les gestes en direction du côté gauche de leur abdomen semblent plus faciles à réaliser. Ils indiquent également que les gestes vers les directions cardinales sont plus faciles à réaliser que ceux vers les directions diagonales. Ces résultats informatifs peuvent guider les choix des concepteurs qui souhaiteraient intégrer de tels gestes dans



une application.

---

## CONCLUSION ET PERSPECTIVES

---

La télévision est une source de communication d'information et de divertissement très populaire. Avec la possibilité récente de connecter les télévisions à Internet, les fournisseurs de contenus offrent toujours plus de nouveaux services et de programmes rendant le téléspectateur de plus en plus actif avec la télévision.

Les télécommandes dotée de toujours plus de boutons ont compliqué l'interaction avec la télévision. Pour limiter ce nombre, les concepteurs proposent des menus hiérarchiques complexes ou de longues listes d'éléments dans lesquels le téléspectateur navigue en appuyant répétitivement sur les touches directionnelles. Ces solutions peinent à répondre correctement aux besoins des téléspectateurs pour naviguer, rechercher, sélectionner du contenu ou saisir de l'information. Pour cette raison, l'amélioration des techniques d'interaction et de ces interfaces pour la télévision et les services qu'elle propose constitue un enjeu scientifique et industriel notable.

Cette partie dresse un résumé des contributions de ce travail de thèse entrepris grâce au soutien du projet *QUAERO*<sup>1</sup>. Nous présentons également quelques perspectives de recherche qu'ont amorcées ces travaux.

### 7.1 CONTRIBUTIONS

Le travail présenté dans cette thèse se compose de trois parties. Après une description des enjeux et du contexte de l'interaction avec la télévision interactive, la première partie s'intéresse à classer les travaux existants pour améliorer les télécommandes. Il en ressort entre autres que la modalité gestuelle permet de focaliser l'attention visuelle sur l'écran de la télévision. Intéressés par cette modalité, nous proposons dans la seconde partie une caractérisation de l'interaction gestuelle qui tente d'unifier l'interaction avec différents types

---

1. <http://www.quaero.org/>

de dispositifs et l'interaction sans dispositif. Enfin, inspirés par cette classification, nous proposons dans la dernière partie plusieurs techniques d'interaction gestuelle suivant deux axes de recherche, selon qu'elles exploitent un dispositif physique ou non.

#### 7.1.1 *Espaces de caractérisation*

Afin d'alimenter la procédure de conception de nouvelles techniques d'interaction pour la télévision interactive, nous avons réalisé un travail préparatoire qui s'est déroulé en deux phases. Tout d'abord, faisant suite au constat des limitations de la télécommande traditionnelle, nous avons dressé une caractérisation de la télécommande basée sur les travaux de la littérature. Cette caractérisation met en avant le type de la télécommande selon qu'elles sont dédiées ou non au contrôle de la télévision et le type d'interaction qui lui est associé. Elle revient également sur les propriétés interactionnelles exploitées dans ces travaux. Parmi les observations qui en découlent, nous remarquons que l'interaction gestuelle pourrait être un moyen de garder l'attention visuelle des téléspectateurs vers l'écran de la télévision. Ceci nous conduit à la seconde phase de notre recherche.

La caractérisation précédente montre la diversité des types de télécommandes. Cependant, plusieurs d'entre elles utilisent des techniques d'interaction gestuelle différentes. Pour guider la conception de nouvelles techniques d'interaction basée sur la modalité gestuelle, nous avons tenté d'unifier l'interaction gestuelle par une caractérisation qui s'appuie sur 5 dimensions :

- *Le type d'interaction* qui définit si l'interaction gestuelle est contrainte par une surface ou non (i.e. main libres). Elle peut également être instrumentée ou non dans les deux cas.
- *La fonction du geste* qui définit si le geste permet d'acquérir de l'information ou de manipuler un objet.
- *L'anatomie du geste* qui décrit d'abord le type de délimiteur du geste utile. Cette dimension décrit également la forme et la cinématique du geste.
- *la combinaison des gestes* qui définit, si elle existe, la combinaison séquentielle ou concurrente de plusieurs gestes.

- *le contexte* qui décrit les contraintes mécaniques, perceptives et sociales de la situation dans laquelle le geste est réalisé.

### 7.1.2 *Techniques d'interaction*

Suite à ce travail, nous avons conçu diverses techniques d'interaction gestuelle permettant d'accéder rapidement aux commandes fréquentes ou aux contenus favoris en évitant d'avoir à naviguer dans des menus hiérarchiques linéaires complexes. Les techniques d'interaction proposées dans le chapitre 5 se basent sur l'utilisation d'une télécommande dotées de capteurs d'orientation en explorant des combinaisons de deux modalités d'interaction différentes (pression de boutons et gestes en l'air). La technique d'interaction présentée dans le chapitre 6 s'abstrait de tout instrument physique et propose de réaliser des gestes sur la surface du corps.

Pour les premières techniques, nous avons d'abord observé des utilisateurs effectuer spontanément des gestes en l'air avec une télécommande pour indiquer une direction. Nous avons ensuite exploré les limites de la discrétisation de l'amplitude du mouvement de roulis, qu'ils n'utilisaient pas, pour augmenter la bande passante interactionnelle. Les résultats montrent que les utilisateurs sont capables de discrétiser l'amplitude du roulis jusqu'à 7 niveaux sans regarder. En considérant les rotations et les boutons de la télécommande, nous avons conçu et évalué des techniques d'interaction qui combinent ces deux types de dimensions. Les résultats montrent que certaines techniques hybrides qui combinent les gestes en l'air et la pression de boutons sont en général très performantes et pour certaines même encore plus que celles basées sur une succession de pressions de boutons. Ces techniques hybrides peuvent compléter les usages habituels qui se basent sur des techniques de pressions de boutons de la télécommande et ainsi augmenter l'expressivité interactionnelle tout en réduisant le nombre de boutons sur le dispositif.

La dernière technique que nous présentons dans ce travail de recherche se base sur des gestes s'appuyant sur la proprioception et réalisés sur la surface du ventre. Dans un premier temps, les spécificités de l'orientation de cette surface nous ont poussé à comprendre la représentation spatiale mentale des utilisateurs quand ils interagissent sans regarder. Cette représentation se base sur trois perspectives différentes exprimées par les utilisateurs : comme si il regarde leur abdomen, comme si il se regarde dans un miroir ou comme si il communique un symbole à une autre personne. Nos résultats montrent que la réalisation des gestes de forme complexe, tels que des chiffres, de-

mande un effort cognitif important et est enclin à une désorientation spatiale importante. En revanche, la représentation est stable pour des gestes de forme simple comme les gestes directionnels. L'étude de l'exécution motrice des gestes directionnels montre que les gestes pour des directions diagonales sont plus difficiles à réaliser que les gestes pour les directions cardinales. Ces résultats sont à la fois utiles pour guider les concepteurs souhaitant intégrer ce type de gestes dans leurs applications ou la conception d'algorithmes pour la reconnaissance des gestes sur cette surface.

## 7.2 PERSPECTIVES

Nous présentons quelques perspectives de travaux futurs à court, moyen et long terme.

### 7.2.1 Perspectives à court terme

Bien qu'exhibant des résultats honorables, nous avons vu que les résultats de l'expérimentation à la section 5.7.2 sur la comparaison de techniques combinant différentes modalités ont été entachés par un algorithme de filtrage de signal trop limité. Afin de mettre à jour les mesures de performance de référence des différentes techniques évaluées, nous souhaitons reconduire une expérimentation avec l'utilisation du *filtre 1 $\epsilon$*  permettant de filtrer le signal avec une réponse très rapide [Casiez et al., 2012].

### 7.2.2 Perspectives à moyen terme

Les techniques d'interaction résultant du chapitre 5 sont inspirées des *Marking Menus* [Kurtenbach et Buxton, 1991]. Ces menus sont connus pour être facilement mémorisables grâce notamment à la pratique continue du traçage des marques directionnelles pour sélectionner un élément [Kurtenbach et Buxton, 1994]. Dans notre expérimentation, nos techniques sont une adaptation des *Marking Menus* avec des gestes en l'air. L'utilisation du tangage et du lacet permettent de sélectionner une direction du menu. Cependant, il n'est pas certain que ces gestes bénéficient des mêmes propriétés de mémorisation que les gestes contraints par une surface comme dans les *Marking Menus* originaux. Par ailleurs, nous avons également proposé d'utiliser des gestes de roulis pour sélectionner un élément dans un menu semi-circulaire. De la même manière, l'efficacité de la mémorisation des éléments du menu par un tel geste reste une question ouverte. Nous souhaitons répondre à ces questions par une expérimentation contrôlée longitudinale en les comparant aux

gestes des *Marking Menus* originaux. Cela permettrait de généraliser ce qui est connu dans le cas de l'interaction sur surface 2D à un autre contexte.

Avec *Belly gestures*, nous avons étudié la représentation mentale spatiale sur l'abdomen lorsque l'utilisateur interagit en s'appuyant uniquement sur la proprioception. Nos résultats montrent que cette représentation dépend de la perspective dans laquelle se projette les utilisateurs. Alors que Lin *et al.* [2011] explorent l'interaction sur le bras en se basant uniquement sur la proprioception, il n'est pas évident de prévoir la représentation mentale spatiale qu'adopteront les utilisateurs, surtout sur une partie du corps mobile qui peut changer d'orientation à tout moment. Nous pensons que la perception de l'orientation de la surface pourrait influencer cette représentation. Alors que des travaux récents de recherche proposent des techniques d'interaction sur différentes parties du corps [Harrison *et al.*, 2010, 2011; Sato *et al.*, 2012a; Yang *et al.*, 2012], nous souhaiterions réaliser ce genre d'expérimentation pour les parties du corps qui sont aisément accessibles par la main, comme l'autre main, les bras ou les cuisses.

Lin *et al.* [2011] proposent d'explorer la division de l'avant-bras pour interagir par brèves pressions en se basant uniquement sur la proprioception. Les résultats de leurs expérimentations montrent d'une part que le retour passif de la peau sur la surface d'interaction améliore la précision des gestes sur le corps et d'autre part que les utilisateurs peuvent discriminer jusqu'à 6 zones distinctes sur l'avant-bras sans le regarder. Les dimensions de la surface de l'abdomen offrent un rapport qui permet d'interagir en deux dimensions avec les deux mains. Cette surface pourrait par exemple être utilisée pour du pointage direct. Cependant, sa résolution lorsque l'utilisateur interagit sans regarder reste inconnue. Similairement à Lin *et al.* [2011], il serait intéressant, dans un premier temps, d'étudier la résolution de cette surface sur une dimension (horizontale et verticale) et de comparer la sensibilité de celle-ci avec la sensibilité de l'avant bras. Cette sensibilité pourrait avoir une influence sur la précision des gestes. A partir de ces résultats, l'étude de la résolution de l'espace en deux dimensions viendrait compléter cette analyse. Notons que la précision de technologie Kinect, utilisée pour nos expérimentations sur l'interaction sur le corps dans cette thèse, n'est pas suffisante pour ce genre d'exploration. Par contre, les systèmes de capture de mouvement, tels que ceux commercialisés par l'entreprise VICON<sup>2</sup>, semblent plus adéquats de par leur grande précision et leur capacité à délimiter l'espace d'interaction abdominale avec des marqueurs.

---

2. <http://www.vicon.com/>

### 7.2.3 Perspectives à long terme

Nous avons vu à la section 6.2.3 que la surface du corps présente de nombreuses caractéristiques intéressantes. Certaines de ces caractéristiques ont déjà été étudiées dans la littérature. Alors que Roudaut *et al.* [2011] ont investigué les gestes de toucher sur les surfaces courbes, Voelker *et al.* [2012] ont tenté de comprendre comment les utilisateurs réalisent des gestes de glissements sur ces mêmes surfaces. Le contrôle du retour haptique procuré par l'adhérence et la texture des surfaces a été exploré pour transmettre de l'information aux doigts [Bau *et al.*, 2010; Bau et Poupyrev, 2012; Casiez *et al.*, 2011]. Le contrôle du niveau de pression a fait également l'objet d'études lorsque le doigt est immobile [Stewart *et al.*, 2010] ou lors d'un geste de glissement [Heo et Lee, 2012]. Toutes ces études ont été réalisées en utilisant des surfaces rigides. Parallèlement, des travaux récents présentent des solutions pour contrôler la rigidité d'une surface [Follmer *et al.*, 2012]. Notons aussi que la surface du corps est réversiblement déformable grâce à son élasticité. Nous pensons que les connaissances déjà acquises dans le cadre des surfaces rigides pourraient être différentes dans le cas des surfaces élastiques. Nous souhaiterions compléter ces connaissances par des études supplémentaires dans le cadre des surfaces élastiques. Enfin, la section 6.2.4.3 fait état de diverses possibilités offertes par l'élasticité d'une surface, comme fournir un guide physique pour les interactions tactiles ou un moyen de sélectionner un mode, que nous souhaiterions examiner.

Pour finir, nous aimerions également apporter une contribution aux méthodes d'évaluation des techniques d'interaction dans le contexte du divertissement. Dans la section 5.7.2, les utilisateurs semblent préférer une des techniques (T&L/Roulis) qui montre un taux de performance en deçà des meilleures. Des résultats similaires ont été relevés dans [Drucker *et al.*, 2002]. La technique pour zapper des séquences dénuées d'intérêt dans un programme TV que proposent les auteurs est jugée qualitativement plus amusante et moins frustrante que toutes les autres techniques à laquelle ils la comparent. Pourtant, c'est également la technique qui demande le plus de temps et le plus d'actions pour réaliser la tâche. La télévision interactive est un outil pour le divertissement et le plaisir d'utilisation est donc un facteur important à prendre en compte lorsque l'on conçoit et évalue une technique d'interaction dans ce contexte. Ce facteur pourrait de plus jouer un rôle essentiel dans l'acceptation des techniques et des dispositifs dans l'environnement domestique [Choi *et al.*, 2003]. De nouvelles méthodes d'évaluation sont alors nécessaires pour répondre aux besoins des utilisateurs.

---

## BIBLIOGRAPHIE

---

- AOKI, R., CHAN, B., IHARA, M., KOBAYASHI, T., KOBAYASHI, M. et KAGAMI, S. (2012). A gesture recognition algorithm for vision-based unicursal gesture interfaces. In *Proceedings of the 10th European conference on Interactive tv and video*, EuroITV '12, pages 53–56, New York, NY, USA. ACM.
- AOKI, R., MAEDA, A., WATANABE, T., KOBAYASHI, M. et ABE, M. (2010). Twist tap : text entry for tv remotes using easy-to-learn wrist motion and key operation. *Consumer Electronics, IEEE Transactions on*, 56(1):161–168.
- BAGLIONI, M., LECOLINET, E. et GUIARD, Y. (2009). Espace de caractérisation des interactions gestuelles physiques sur dispositifs mobiles. In *Proceedings of the 21st International Conference on Association Francophone d'Interaction Homme-Machine*, IHM '09, pages 203–212, New York, NY, USA. ACM.
- BAGLIONI, M., MALACRIA, S., LECOLINET, E. et GUIARD, Y. (2011). Flick-and-brake : finger control over inertial/sustained scroll motion. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 2281–2286, New York, NY, USA. ACM.
- BAILLY, G., LECOLINET, E. et GUIARD, Y. (2010). Finger-count & radial-stroke shortcuts : 2 techniques for augmenting linear menus on multi-touch surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 591–594, New York, NY, USA. ACM.
- BAILLY, G., LECOLINET, E. et NIGAY, L. (2008). Flower menus : a new type of marking menu with large menu breadth, within groups and efficient expert mode memorization. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '08, pages 15–22, New York, NY, USA. ACM.
- BAILLY, G., MÜLLER, J., ROHS, M., WIGDOR, D. et KRATZ, S. (2012). Shoesense : a new perspective on gestural interaction and wearable applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1239–1248, New York, NY, USA. ACM.
- BAILLY, G., WALTER, R., MÜLLER, J., NING, T. et LECOLINET, E. (2011). Comparing free hand menu techniques for distant displays using linear, marking and finger-count menus. In CAMPOS, P., GRAHAM, N., JORGE, J., NUNES, N., PALANQUE, P. et WINCKLER, M., éditeurs : *Human-Computer Interaction –*



*INTERACT 2011*, volume 6947 de *Lecture Notes in Computer Science*, pages 248–262. Springer Berlin Heidelberg.

- BALCHANDRAN, R., EPSTEIN, M. E., POTAMIANOS, G. et SEREDI, L. (2008). A multi-modal spoken dialog system for interactive tv. In *Proceedings of the 10th international conference on Multimodal interfaces, ICMI '08*, pages 191–192, New York, NY, USA. ACM.
- BALLAGAS, R., ROHS, M. et SHERIDAN, J. G. (2005). Sweep and point and shoot : phonecam-based interactions for large public displays. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems, CHI EA '05*, pages 1200–1203, New York, NY, USA. ACM.
- BALLENDAT, T., MARQUARDT, N. et GREENBERG, S. (2010). Proxemic interaction : designing for a proximity and orientation-aware environment. In *ACM International Conference on Interactive Tabletops and Surfaces, ITS '10*, pages 121–130, New York, NY, USA. ACM.
- BARTLETT, J. (2000). Rock 'n' scroll is here to stay [user interface]. *Computer Graphics and Applications, IEEE*, 20(3):40–45.
- BAU, O. et POUPYREV, I. (2012). Revel : tactile feedback technology for augmented reality. *ACM Trans. Graph.*, 31(4):89 :1–89 :11.
- BAU, O., POUPYREV, I., ISRAR, A. et HARRISON, C. (2010). Teslatouch : electrovibration for touch surfaces. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology, UIST '10*, pages 283–292, New York, NY, USA. ACM.
- BAUDEL, T. et BEAUDOUIN-LAFON, M. (1993). Charade : remote control of objects using free-hand gestures. *Commun. ACM*, 36(7):28–35.
- BEAUDOUIN-LAFON, M. (2000). Instrumental interaction : An interaction model for designing post-wimp user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '00*, pages 446–453, New York, NY, USA. ACM.
- BERGLUND, A., BERGLUND, E., LARSSON, A. et BANG, M. (2006). Paper remote : an augmented television guide and remote control. *Universal Access in the Information Society*, 4(4):300–327.
- BERGLUND, A. et JOHANSSON, P. (2004). Using speech and dialogue for interactive tv navigation. *Universal Access in the Information Society*, 3(3-4):224–238.
- BERNHaupt, R., OBRIST, M., WEISS, A., BECK, E. et TSCHELIGI, M. (2008). Trends in the living room and beyond : results from ethnographic studies using creative and playful probing. *Comput. Entertain.*, 6(1):5 :1–5 :23.

- BIET, M., GIRAUD, F. et LEMAIRE-SEMAIL, B. (2007). Squeeze film effect for the design of an ultrasonic tactile plate. *Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on*, 54(12):2678–2688.
- BLACK, A., BAYLEY, O., BURNS, C., KUULUVAINEN, I. et STODDARD, J. (1994). Keeping viewers in the picture : real-world usability procedures in the development of a television control interface. In *Conference Companion on Human Factors in Computing Systems, CHI '94*, pages 243–244, New York, NY, USA. ACM.
- BLOCK, F., SCHMIDT, A., VILLAR, N. et GELLERSEN, H. (2004). Towards a playful user interface for home entertainment systems. In MARKOPOULOS, P., EGGEN, B., AARTS, E. et CROWLEY, J., éditeurs : *Ambient Intelligence*, volume 3295 de *Lecture Notes in Computer Science*, pages 207–217. Springer Berlin Heidelberg.
- BORING, S., JURMU, M. et BUTZ, A. (2009). Scroll, tilt or move it : using mobile phones to continuously control pointers on large public displays. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group : Design : Open 24/7, OZCHI '09*, pages 161–168, New York, NY, USA. ACM.
- CADOZ, C. (1994). Le geste canal de communication homme/machine. la communication 'instrumentale'. *Technique et science informatiques*, 13(1):31–61.
- CASIEZ, G., ROUSSEL, N., VANBELLEGHEM, R. et GIRAUD, F. (2011). Surfpad : riding towards targets on a squeeze film effect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, pages 2491–2500, New York, NY, USA. ACM.
- CASIEZ, G., ROUSSEL, N. et VOGEL, D. (2012). 1 &#364; filter : a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, pages 2527–2530, New York, NY, USA. ACM.
- CESAR, P., BULTERMAN, D. et JANSEN, J. (2009). Leveraging user impact : an architecture for secondary screens usage in interactive television. *Multimedia Systems*, 15(3):127–142.
- CESAR, P., CHORIANOPOULOS, K. et JENSEN, J. F. (2008). Social television and user interaction. *Comput. Entertain.*, 6(1):4 :1–4 :10.
- CHOI, H., CHOI, M., KIM, J. et YU, H. (2003). An empirical study on the adoption of information appliances with a focus on interactive tv. *Telematics and Informatics*, 20(2):161 – 183.

- CHOI, S., HAN, J., LEE, G., LEE, N. et LEE, W. (2011). Remotetouch : touch-screen-like interaction in the tv viewing environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, pages 393–402, New York, NY, USA. ACM.
- CHORIANOPOULOS, K. (2008). User interface design principles for interactive television applications. *International Journal of Human-Computer Interaction*, 24(6):556–573.
- CHORIANOPOULOS, K. et SPINELLIS, D. (2006). User interface evaluation of interactive tv : a media studies perspective. *Universal Access in the Information Society*, 5(2):209–218.
- COOPER, W. (2008). The interactive television user experience so far. In *Proceedings of the 1st international conference on Designing interactive user experiences for TV and video, UXTV '08*, pages 133–142, New York, NY, USA. ACM.
- COVER, T. et HART, P. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27.
- CRUICKSHANK, L., TSEKLEVES, E., WHITHAM, R., HILL, A. et KONDO, K. (2007). Making interactive tv easier to use : Interface design for a second screen approach. *The Design Journal*, 10(3):41–53.
- DRUCKER, S. M., GLATZER, A., DE MAR, S. et WONG, C. (2002). Smartskip : consumer level browsing and skipping of digital video content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '02*, pages 219–226, New York, NY, USA. ACM.
- EFRON, D. (1941). *Gesture and environment*. King's Crown Press.
- ENNS, N. R. N. et MACKENZIE, I. S. (1998). Touchpad-based remote control devices. In *CHI 98 Conference Summary on Human Factors in Computing Systems, CHI '98*, pages 229–230, New York, NY, USA. ACM.
- FERSCHA, A., VOGL, S., EMSENHUBER, B. et WALLY, B. (2007). Physical shortcuts for media remote controls. In *Proceedings of the 2nd international conference on INtelligent TEchnologies for interactive enterTAINment, INTETAIN '08*, pages 9 :1–9 :8, ICST, Brussels, Belgium, Belgium. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- FITZMAURICE, G. W., ZHAI, S. et CHIGNELL, M. H. (1993). Virtual reality for palmtop computers. *ACM Trans. Inf. Syst.*, 11(3):197–218.
- FOLEY, J., WALLACE, V. et CHAN, P. (1984). The human factors of computer graphics interaction techniques. *Computer Graphics and Applications, IEEE*, 4(11):13–48.

- FOLLMER, S., LEITHINGER, D., OLWAL, A., CHENG, N. et ISHII, H. (2012). Jamming user interfaces : programmable particle stiffness and sensing for malleable and shape-changing devices. In *Proceedings of the 25th annual ACM symposium on User interface software and technology, UIST '12*, pages 519–528, New York, NY, USA. ACM.
- FORLINES, C., WIGDOR, D., SHEN, C. et BALAKRISHNAN, R. (2007). Direct-touch vs. mouse input for tabletop displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07*, pages 647–656, New York, NY, USA. ACM.
- FREEMAN, D., BENKO, H., MORRIS, M. R. et WIGDOR, D. (2009). Shadowguides : visualizations for in-situ learning of multi-touch and whole-hand gestures. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, ITS '09*, pages 165–172, New York, NY, USA. ACM.
- FREEMAN T., W. et CRAIG D., W. (1995). Television control by hand gestures. In *IEEE International Workshop on Automatic Face- and Gesture-Recognition*. IEEE.
- GALLACE, A. et SPENCE, C. (2010). The science of interpersonal touch : An overview. *Neuroscience & Biobehavioral Reviews*, 34(2):246 – 259. Touch, Temperature, Pain/Itch and Pleasure.
- GAWLINSKI, M. (2003). *Interactive Television Production*. Taylor & Francis.
- GEMPERLE, F., KASABACH, C., STIVORIC, J., BAUER, M. et MARTIN, R. (1998). Design for wearability. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on*, pages 116–122.
- GENTAZ, E., BAUD-BOVY, G. et LUYAT, M. (2008). The haptic perception of spatial orientations. *Experimental Brain Research*, 187(3):331–348.
- GOTO, J., KIM, Y.-B., MIYAZAKI, M., KOMINE, K. et URATANI, N. (2003). A spoken dialogue interface for tv operations based on data collected by using woz method. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 2, ACL '03*, pages 101–104, Stroudsburg, PA, USA. Association for Computational Linguistics.
- GREENBERG, S., MARQUARDT, N., BALLENDAT, T., DIAZ-MARINO, R. et WANG, M. (2011). Proxemic interactions : the new ubicomp ? *interactions*, 18(1):42–50.
- GUEDDANA, S. et ROUSSEL, N. (2009). Pêle-mêle, une étude de la communication multi-échelles. In *Proceedings of the 21st International Conference on Association Francophone d'Interaction Homme-Machine, IHM '09*, pages 273–282, New York, NY, USA. ACM.

- GURRIN, C., LEE, H., CAPRANI, N., ZHANG, Z., O'CONNOR, N. et CARTHY, D. (2010). Browsing large personal multimedia archives in a lean-back environment. In BOLL, S., TIAN, Q., ZHANG, L., ZHANG, Z. et CHEN, Y.-P., éditeurs : *Advances in Multimedia Modeling*, volume 5916 de *Lecture Notes in Computer Science*, pages 98–109. Springer Berlin Heidelberg.
- GUSTAFSON, S., HOLZ, C. et BAUDISCH, P. (2011). Imaginary phone : learning imaginary interfaces by transferring spatial memory from a familiar device. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 283–292, New York, NY, USA. ACM.
- HARRISON, C., BENKO, H. et WILSON, A. D. (2011). Omnitouch : wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 441–450, New York, NY, USA. ACM.
- HARRISON, C., TAN, D. et MORRIS, D. (2010). Skinput : appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 453–462, New York, NY, USA. ACM.
- HEO, S. et LEE, G. (2012). Forcedrag : using pressure as a touch input modifier. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, OzCHI '12, pages 204–207, New York, NY, USA. ACM.
- HESS, J., KÜSTERMANN, G. et PIPEK, V. (2008). Premote : a user customizable remote control. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 3279–3284, New York, NY, USA. ACM.
- HOGAN, N. (1985). The mechanics of multi-joint posture and movement control. *Biological Cybernetics*, 52(5):315–331.
- HOLMES, M. E., JOSEPHSON, S. et CARNEY, R. E. (2012). Visual attention to television programs with a second-screen application. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, pages 397–400, New York, NY, USA. ACM.
- HOLZ, C., GROSSMAN, T., FITZMAURICE, G. et AGUR, A. (2012). Implanted user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 503–512, New York, NY, USA. ACM.
- HUDSON, S. E., HARRISON, C., HARRISON, B. L. et LAMARCA, A. (2010). Whack gestures : inexact and inattentive interaction with mobile devices. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, TEI '10, pages 109–112, New York, NY, USA. ACM.

- HUMMELS, C. et STAPPERS, P. (1998). Meaningful gestures for human computer interaction : beyond hand postures. *In Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pages 591–596.
- IBRAHIM, A., LUNDBERG, J. et JOHANSSON, J. (2001). Speech enhanced remote control for media terminal. *In Proceedings of the 7th European Conference on Speech Communication and Technology, EUROSPEECH 2001*, pages 2685–2688.
- INGMARSSON, M., DINKA, D. et ZHAI, S. (2004). Tnt : a numeric keypad based text input method. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '04*, pages 639–646, New York, NY, USA. ACM.
- ISHIYAMA, K. et YANO, S. (2000). A study of characteristics of pointing devices for television operation. *In Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, volume 2, pages 1307–1312 vol.2.
- JACOB, R. J. K., SIBERT, L. E., MCFARLANE, D. C. et MULLEN, Jr., M. P. (1994). Integrality and separability of input devices. *ACM Trans. Comput.-Hum. Interact.*, 1(1):3–26.
- JENSEN, J. F. (2005). Interactive television : new genres, new format, new content. *In Proceedings of the second Australasian conference on Interactive entertainment, IE '05*, pages 89–96, Sydney, Australia, Australia. Creativity & Cognition Studios Press.
- JENSEN, J. F. (2008). Interactive television - a brief media history. *In TSCHELIGI, M., OBRIST, M. et LUGMAYR, A., éditeurs : Changing Television Environments*, volume 5066 de *Lecture Notes in Computer Science*, pages 1–10. Springer Berlin Heidelberg.
- JONES, E., ALEXANDER, J., ANDREOU, A., IRANI, P. et SUBRAMANIAN, S. (2010). Gestext : accelerometer-based gestural text-entry systems. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10*, pages 2173–2182, New York, NY, USA. ACM.
- KALMAN, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, 82(Series D):35–45.
- KARAM, M. et m. c. SCHRAEFEL (2005). A taxonomy of gestures in human computer interactions. Technical report, University of Southampton.
- KARRER, T., WITTENHAGEN, M., LICHTSCHLAG, L., HELLER, F. et BORCHERS, J. (2011). Pinstripe : eyes-free continuous input on interactive clothing. *In*



- Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, pages 1313–1322, New York, NY, USA. ACM.
- KENDON, A. (1980). Gesticulation and speech : Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication*, 25:207–227.
- KIMMAN, F., WEDA, H., van den HOVEN, E., de ZEEUW, T. et LUITJENS, S. (2011). Spinning in control : design exploration of a cross-device remote. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction, TEI '11*, pages 189–192, New York, NY, USA. ACM.
- KNOBLOCH, S. et ZILLMANN, D. (2002). Mood management via the digital jukebox. *Journal of Communication*, 52(2):351–366.
- KOBAYASHI, K., NAKAGAWA, Y., YAMADA, S., NAKAGAWA, S. et SAITO, Y. (2009). Rebo : A remote control with strokes. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 751–756.
- KURTENBACH, G. et BUXTON, W. (1991). Issues in combining marking and direct manipulation techniques. In *Proceedings of the 4th annual ACM symposium on User interface software and technology, UIST '91*, pages 137–144, New York, NY, USA. ACM.
- KURTENBACH, G. et BUXTON, W. (1994). User learning and performance with marking menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '94*, pages 258–264, New York, NY, USA. ACM.
- KURTENBACH, G. P., SELLEN, A. J. et BUXTON, W. A. (1993). An empirical evaluation of some articulatory and cognitive aspects of marking menus. *Human–Computer Interaction*, 8(1):1–23.
- LEE, C.-H. J., CHANG, C., CHUNG, H., DICKIE, C. et SELKER, T. (2007). Emotionally reactive television. In *Proceedings of the 12th international conference on Intelligent user interfaces, IUI '07*, pages 329–332, New York, NY, USA. ACM.
- LEE, D.-W., MOOK LIM, J., SUNWOO, J., CHO, I.-Y. et LEE, C.-H. (2009). Actual remote control : a universal remote control using hand motions on a virtual menu. *Consumer Electronics, IEEE Transactions on*, 55(3):1439–1446.
- LEPINSKI, G. J., GROSSMAN, T. et FITZMAURICE, G. (2010). The design and evaluation of multitouch marking menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10*, pages 2233–2242, New York, NY, USA. ACM.

- LIMA, J. V. d. et NEVADO, R. A. d. (2007). Common paper as an interface for digital tv. In *Multimedia Workshops, 2007. ISMW '07. Ninth IEEE International Symposium on*, pages 357–362.
- LIN, S.-Y., SU, C.-H., CHENG, K.-Y., LIANG, R.-H., KUO, T.-H. et CHEN, B.-Y. (2011). Pub - point upon body : exploring eyes-free interaction and methods on an arm. In *Proceedings of the 24th annual ACM symposium on User interface software and technology, UIST '11*, pages 481–488, New York, NY, USA. ACM.
- LOCLAIR, C., GUSTAFSON, S. et BAUDISCH, P. (2010). Pinchwatch : a wearable device for one-handed microinteractions. In *MobileHCI Workshop on Ensembles of On-Body Devices*, page 4 Pages.
- LOPES, P., BUTZMANN, L. et BAUDISCH, P. (2013). Muscle-propelled force feedback : bringing force feedback to mobile devices using electrical stimulation. In *Proceedings of the 4th Augmented Human International Conference, AH '13*, pages 231–232, New York, NY, USA. ACM.
- LÜ, H. et LI, Y. (2011). Gesture avatar : a technique for operating mobile user interfaces using gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, pages 207–216, New York, NY, USA. ACM.
- MACKENZIE, I. S. et JUSOH, S. (2001). An evaluation of two input devices for remote pointing. In *Proceedings of the 8th IFIP International Conference on Engineering for Human-Computer Interaction, EHCI '01*, pages 235–250, London, UK, UK. Springer-Verlag.
- MALACRIA, S. et LECOLINET, E. (2008). Espace de caractérisation du stylo numérique. In *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine, IHM '08*, pages 177–184, New York, NY, USA. ACM.
- MALACRIA, S., LECOLINET, E. et GUIARD, Y. (2010). Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces : the cyclostar approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10*, pages 2615–2624, New York, NY, USA. ACM.
- MARSHALL, D., FOSTER, J. C. et JACK, M. A. (2001). User performance and attitude towards schemes for alphanumeric data entry using restricted input devices. *Behaviour & Information Technology*, 20(3):167–188.
- MCNEILL, D. (1992). *Hand and mind : What gestures reveal about thought*. University of Chicago Press.



- MEERBEEK, B., BINGLEY, P., RIJNEN, W. et van den HOVEN, E. (2010). Pipet : a design concept supporting photo sharing. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction : Extending Boundaries*, NordiCHI '10, pages 335–342, New York, NY, USA. ACM.
- MERGNER, T. et ROSEMEIER, T. (1998). Interaction of vestibular, somatosensory and visual signals for postural control and motion perception under terrestrial and microgravity conditions—a conceptual model. *Brain research. Brain research reviews*, 28(1-2):118–35.
- MIRLACHER, T., PIRKER, M., BERNHAUPT, R., FISCHER, T., SCHWAIGER, D., WILFINGER, D. et TSCHELIGI, M. (2010). Interactive simplicity for itv : minimizing keys for navigating content. In *Proceedings of the 8th international interactive conference on Interactive TV&#38;Video*, EuroITV '10, pages 137–140, New York, NY, USA. ACM.
- NACK, F., SCHIPHORST, T., OBRENOVIC, Z., KAUWATJOE, M., de BAKKER, S., ROSILLIO, A. P. et AROYO, L. (2007). Pillows as adaptive interfaces in ambient environments. In *Proceedings of the international workshop on Human-centered multimedia*, HCM '07, pages 3–12, New York, NY, USA. ACM.
- NAKATSUMA, K., SHINODA, H., MAKINO, Y., SATO, K. et MAENO, T. (2011). Touch interface on back of the hand. In *ACM SIGGRAPH 2011 Emerging Technologies*, SIGGRAPH '11, pages 19 :1–19 :1, New York, NY, USA. ACM.
- NANCEL, M., WAGNER, J., PIETRIGA, E., CHAPUIS, O. et MACKAY, W. (2011). Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 177–186, New York, NY, USA. ACM.
- NI, T., MCMAHAN, R. et BOWMAN, D. (2008). Tech-note : rapmenu : Remote menu selection using freehand gestural input. In *3D User Interfaces, 2008. 3DUI 2008. IEEE Symposium on*, pages 55–58.
- OAKLEY, I. et O'MODHRAIN, S. (2005). Tilt to scroll : Evaluating a motion based vibrotactile mobile interface. *World Haptics Conference*, 0:40–49.
- OAKLEY, I. et PARK, J. (2009). Motion marking menus : An eyes-free approach to motion input for handheld devices. *International Journal of Human-Computer Studies*, 67(6):515 – 532.
- OAKLEY, I. et PARK, J.-S. (2007). Designing eyes-free interaction. In OAKLEY, I. et BREWSTER, S., éditeurs : *Haptic and Audio Interaction Design*, volume 4813 de *Lecture Notes in Computer Science*, pages 121–132. Springer Berlin Heidelberg.

- OLSEN, D. R., PARTRIDGE, B. et LYNN, S. (2010). Time warp sports for internet television. *ACM Trans. Comput.-Hum. Interact.*, 17(4):16 :1–16 :37.
- O'MODHRAIN, S. et OAKLEY, I. (2003). Touch tv : Adding feeling to broadcast media. In *in proceedings of the European Conference on Interactive Television : from Viewers to Actors*, pages 41–47.
- O'MODHRAIN, S. et OAKLEY, I. (2004). Adding interactivity : active touch in broadcast media. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS '04. Proceedings. 12th International Symposium on*, pages 293–294.
- PARSONS, L. M. et SHIMOJO, S. (1987). Perceived spatial organization of cutaneous patterns on surfaces of the human body in various positions. *Journal of Experimental Psychology : Human Perception and Performance*, 13(3):488–504.
- PARTRIDGE, K., CHATTERJEE, S., SAZAWAL, V., BORRIELLO, G. et WANT, R. (2002). Tilttype : accelerometer-supported text entry for very small devices. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, UIST '02, pages 201–204, New York, NY, USA. ACM.
- PEDERSEN, E. W. et HORNBAEK, K. (2012). An experimental comparison of touch interaction on vertical and horizontal surfaces. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction : Making Sense Through Design*, NordiCHI '12, pages 370–379, New York, NY, USA. ACM.
- PERRAULT, S. T., BAILLY, G., GUIARD, Y. et LECOLINET, E. (2011). Promesses et contraintes de la joaillerie numérique interactive : un aperçu de l'état de l'art. In *23rd French Speaking Conference on Human-Computer Interaction, IHM '11*, pages 14 :1–14 :4, New York, NY, USA. ACM.
- PIRKER, M., BERNHAUPT, R. et MIRLACHER, T. (2010). Investigating usability and user experience as possible entry barriers for touch interaction in the living room. In *Proceedings of the 8th international interactive conference on Interactive TV&#38;Video*, EuroITV '10, pages 145–154, New York, NY, USA. ACM.
- PROSKE, U. et GANDEVIA, S. C. (2012). The proprioceptive senses : Their roles in signaling body shape, body position and movement, and muscle force. *Physiological Reviews*, 92(4):1651–1697.
- QUEK, F., MCNEILL, D., BRYLL, R., DUNCAN, S., MA, X.-F., KIRBAS, C., MCCULLOUGH, K. E. et ANSARI, R. (2002). Multimodal human discourse : gesture and speech. *ACM Trans. Comput.-Hum. Interact.*, 9(3):171–193.

- RAHMAN, M., GUSTAFSON, S., IRANI, P. et SUBRAMANIAN, S. (2009). Tilt techniques : investigating the dexterity of wrist-based input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1943–1952, New York, NY, USA. ACM.
- RAMOS, G. A. et BALAKRISHNAN, R. (2007). Pressure marks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1375–1384, New York, NY, USA. ACM.
- RASKIN, J. (2000). *The Humane Interface : New Directions for Designing Interactive Systems*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA.
- REKIMOTO, J. (1996). Tilting operations for small screen interfaces. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, UIST '96, pages 167–168, New York, NY, USA. ACM.
- REN, G. et O'NEILL, E. (2012). 3d marking menu selection with freehand gestures. In *3D User Interfaces (3DUI), 2012 IEEE Symposium on*, pages 61–68.
- RICO, J. et BREWSTER, S. (2010). Usable gestures for mobile interfaces : evaluating social acceptability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 887–896, New York, NY, USA. ACM.
- ROBERTSON, S., WHARTON, C., ASHWORTH, C. et FRANZKE, M. (1996). Dual device user interface design : Pdas and interactive television. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, pages 79–86, New York, NY, USA. ACM.
- ROUDAUT, A., LECOLINET, E. et GUIARD, Y. (2009). Microrolls : expanding touch-screen input vocabulary by distinguishing rolls vs. slides of the thumb. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 927–936, New York, NY, USA. ACM.
- ROUDAUT, A., POHL, H. et BAUDISCH, P. (2011). Touch input on curved surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1011–1020, New York, NY, USA. ACM.
- ROUDAUT, A., RAU, A., STERZ, C., PLAUTH, M., LOPES, P. et BAUDISCH, P. (2013). Gesture output : eyes-free output using a force feedback touch surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2547–2556, New York, NY, USA. ACM.

- RUIZ, J. et LI, Y. (2011). Doubleflip : a motion gesture delimiter for mobile interaction. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2717–2720, New York, NY, USA. ACM.
- RUIZ, J., LI, Y. et LANK, E. (2011). User-defined motion gestures for mobile interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 197–206, New York, NY, USA. ACM.
- SATO, M., POUPYREV, I. et HARRISON, C. (2012a). Touch : enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 483–492, New York, NY, USA. ACM.
- SATO, T., TAKAHASHI, N., MATOBA, Y. et KOIKE, H. (2012b). Interactive surface that have dynamic softness control. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '12, pages 796–797, New York, NY, USA. ACM.
- SAZAWAL, V., WANT, R. et BORRIELLO, G. (2002). The unigesture approach one-handed text entry for small devices. In PATERNÒ, F., éditeur : *Human Computer Interaction with Mobile Devices*, volume 2411 de *Lecture Notes in Computer Science*, pages 256–270. Springer Berlin Heidelberg.
- SCHMIDT, A., HOLLEIS, P. et KRANZ, M. (2004). Sensor-Virrig - A Balance Cushion as Controller. *Workshop Playing with Sensors at UbiComp 2004*.
- SCODITTI, A., BLANCH, R. et COUTAZ, J. (2011). A novel taxonomy for gestural interaction techniques based on accelerometers. In *Proceedings of the 16th international conference on Intelligent user interfaces*, IUI '11, pages 63–72, New York, NY, USA. ACM.
- SERINO, A. et HAGGARD, P. (2010). Touch and the body. *Neuroscience & Biobehavioral Reviews*, 34(2):224 – 236. <ce :title>Touch, Temperature, Pain/Itch and Pleasure</ce :title>.
- SHANNON, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–.
- SHOEMAKER, G., TSUKITANI, T., KITAMURA, Y. et BOOTH, K. S. (2010). Body-centric interaction techniques for very large wall displays. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction : Extending Boundaries*, NordiCHI '10, pages 463–472, New York, NY, USA. ACM.
- STEWART, C., ROHS, M., KRATZ, S. et ESSL, G. (2010). Characteristics of pressure-based input for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 801–810, New York, NY, USA. ACM.

- SVENSSON, M. S. et SOKOLER, T. (2008). Ticket-to-talk-television : designing for the circumstantial nature of everyday social interaction. In *Proceedings of the 5th Nordic conference on Human-computer interaction : building bridges*, NordiCHI '08, pages 334–343, New York, NY, USA. ACM.
- TAHIR, M., BAILLY, G., LECOLINET, E. et MOURET, G. (2008). Tactimote : a tactile remote control for navigating in long lists. In *Proceedings of the 10th international conference on Multimodal interfaces*, ICMI '08, pages 285–288, New York, NY, USA. ACM.
- TSEKLEVES, E., CRUICKSHANK, L., HILL, A., KONDO, K. et WHITHAM, R. (2007). Interacting with digital media at home via a second screen. In *Multimedia Workshops, 2007. ISMW '07. Ninth IEEE International Symposium on*, pages 201–206.
- TSEKLEVES, E., WHITHAM, R., KONDO, K. et HILL, A. (2011). Investigating media use and the television user experience in the home. *Entertainment Computing*, 2(3):151 – 161. <ce :title>User experiences in the new {TV} landscape</ce :title>.
- TSENG, M. et KOLKO, J. (2005). Triba : a cable television retrieval & awareness system. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1152–1153, New York, NY, USA. ACM.
- TURUNEN, M., KALLINEN, A., SÀNCHEZ, I., RIEKKI, J., HELLA, J., OLSSON, T., MELTO, A., RAJANIEMI, J.-P., HAKULINEN, J., MÄKINEN, E., VALKAMA, P., MIETTINEN, T., PYYKKÖNEN, M., SALORANTA, T., GILMAN, E. et RAISAMO, R. (2009). Multimodal interaction with speech and physical touch interface in a media center application. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology*, ACE '09, pages 19–26, New York, NY, USA. ACM.
- UGARTE, A., GARCÍA, I., ORTIZ, A. et OYARZUN, D. (2007). User interfaces based on 3d avatars for interactive television. In CESAR, P., CHORIANOPoulos, K. et JENSEN, J., éditeurs : *Interactive TV : a Shared Experience*, volume 4471 de *Lecture Notes in Computer Science*, pages 107–115. Springer Berlin Heidelberg.
- VATAVU, R.-D. (2012). User-defined gestures for free-hand tv control. In *Proceedings of the 10th European conference on Interactive tv and video*, EuroITV '12, pages 45–48, New York, NY, USA. ACM.
- VATAVU, R.-D. et PENTIUC, S.-G. (2008). Interactive coffee tables : Interfacing tv within an intuitive, fun and shared experience. In TSCHELIGI, M., OBRIST, M. et LUGMAYR, A., éditeurs : *Changing Television Environments*, volume 5066

- de *Lecture Notes in Computer Science*, pages 183–187. Springer Berlin Heidelberg.
- VILLAR, N. et HODGES, S. (2010). The peppermill : a human-powered user interface device. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction, TEI '10*, pages 29–32, New York, NY, USA. ACM.
- VOELKER, S., SUTTER, C., WANG, L. et BORCHERS, J. (2012). Understanding flicking on curved surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, pages 189–198, New York, NY, USA. ACM.
- WAGNER, J., NANCEL, M., GUSTAFSON, S. G., HUOT, S. et MACKAY, W. E. (2013). Body-centric design space for multi-surface interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13*, pages 1299–1308, New York, NY, USA. ACM.
- WEXELBLAT, A. D. (1994). *A feature-based approach to continuous-gesture analysis*. Thèse de doctorat, Massachusetts Institute of Technology.
- WIGDOR, D. et BALAKRISHNAN, R. (2003). Tilttext : using tilt for text input to mobile phones. In *Proceedings of the 16th annual ACM symposium on User interface software and technology, UIST '03*, pages 81–90, New York, NY, USA. ACM.
- WILSON, A., BOBICK, A. et CASSELL, J. (1996). Recovering the temporal structure of natural gesture. In *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, pages 66–71.
- WILSON, A. et SHAFER, S. (2003). Xwand : Ui for intelligent spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '03*, pages 545–552, New York, NY, USA. ACM.
- WINGRAVE, C., WILLIAMSON, B., VARCHOLIK, P. D., ROSE, J., MILLER, A., CHARBONNEAU, E., BOTT, J. et LAVIOLA, J. (2010). The wiimote and beyond : Spatially convenient devices for 3d user interfaces. *Computer Graphics and Applications, IEEE*, 30(2):71–85.
- WITTENBURG, K., LANNING, T., SCHWENKE, D., SHUBIN, H. et VETRO, A. (2006). The prospects for unrestricted speech input for tv content search. In *Proceedings of the working conference on Advanced visual interfaces, AVI '06*, pages 352–359, New York, NY, USA. ACM.
- WOBBROCK, J. O., MORRIS, M. R. et WILSON, A. D. (2009). User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on*



- Human Factors in Computing Systems*, CHI '09, pages 1083–1092, New York, NY, USA. ACM.
- WOBBOCK, J. O., MYERS, B. A. et AUNG, H. H. (2008). The performance of hand postures in front- and back-of-device interaction for mobile computing. *International Journal of Human-Computer Studies*, 66(12):857 – 875.
- WU, M. et BALAKRISHNAN, R. (2003). Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, UIST '03, pages 193–202, New York, NY, USA. ACM.
- YANG, X.-D., GROSSMAN, T., WIGDOR, D. et FITZMAURICE, G. (2012). Magic finger : always-available input through finger instrumentation. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, pages 147–156, New York, NY, USA. ACM.
- YOO, B., HAN, J.-J., CHOI, C., RYU, H.-s., PARK, D. S. et KIM, C. Y. (2011). 3d remote interface for smart displays. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 551–560, New York, NY, USA. ACM.
- ZHANG, B., CHEN, Y., QIAN, Y. et WANG, X. (2011). A ring-shaped interactive device for large remote display and mobile device control. In *Proceedings of the 13th international conference on Ubiquitous computing*, UbiComp '11, pages 473–474, New York, NY, USA. ACM.
- ZHAO, S., AGRAWALA, M. et HINCKLEY, K. (2006). Zone and polygon menus : using relative position to increase the breadth of multi-stroke marking menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1077–1086, New York, NY, USA. ACM.
- ZHAO, S. et BALAKRISHNAN, R. (2004). Simple vs. compound mark hierarchical marking menus. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, UIST '04, pages 33–42, New York, NY, USA. ACM.
- ZIGELBAUM, J., BROWNING, A., LEITHINGER, D., BAU, O. et ISHII, H. (2010). g-stalt : a chirocentric, spatiotemporal, and telekinetic gestural interface. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, TEI '10, pages 261–264, New York, NY, USA. ACM.

---

## PUBLICATIONS PERSONNELLES

---

- BAILLY, G., VO, D.-B., LECOLINET, E. et GUIARD, Y. (2011). Gesture-aware remote controls : guidelines and interaction technique. In *Proceedings of the 13th international conference on multimodal interfaces, ICMI '11*, pages 263–270, New York, NY, USA. ACM.
- FABIEN, A., ARTIGNAN, G., CORDEIL, M., HOARAU, R., VO, D.-B. et al. (2010). Exploration et gestion des connaissances des ressources en ihm. In *IHM'10 : 22ème Conférence Francophone sur l'Interaction Homme-Machine*.
- VO, D.-B., BAILLY, G., LECOLINET, E. et GUIARD, Y. (2011). Un espace de caractérisation de la télécommande dans le contexte de la télévision interactive. In *23rd French Speaking Conference on Human-Computer Interaction, IHM '11*, pages 17 :1–17 :8, New York, NY, USA. ACM.
- VO, D.-B. et LECOLINET, E. (2013). Belly gestures : Body centric gestures on the abdomen. In *En cours de soumission*.
- VO, D.-B. et WINCKLER, M. (2009). Piaff : un outil d'aide à la saisie d'informations personnelles pour les formulaires électroniques. In *Proceedings of the 21st International Conference on Association Francophone d'Interaction Homme-Machine, IHM '09*, pages 355–358, New York, NY, USA. ACM.
- WINCKLER, M., GAITS, V., VO, D.-B., SERGIO, F. et ROSSI, G. (2011). An approach and tool support for assisting users to fill-in web forms with personal information. In *Proceedings of the 29th ACM international conference on Design of communication, SIGDOC '11*, pages 195–202, New York, NY, USA. ACM.